The Zweitstimme Model: A Dynamic Forecast of the 2021 German Federal Election

Thomas Gschwend, *University of Mannheim, Germany*
Klara Müller, *University of Mannheim, Germany*
Simon Munzert, *The Hertie School, Germany*
Marcel Neunhoeffer, *Ludwig-Maximilians-University Munich, Germany*
Lukas F. Stoetzer, *Humboldt University of Berlin, Germany*

When German citizens head to the polling booths on Sunday, September 26, 2021, no party is expected to gain an outright majority in the Bundestag. Given that several parties are likely to gain representation, only a coalition government will be able to secure a majority of seats. Which parties will gain enough seats and are likely to be in the position to sign a coalition agreement? This election also will determine who will follow Angela Merkel as the new chancellor. Our forecasting project developed the "Zweitstimme" model (i.e., the term for the party vote that Germans cast on Election Day), which performed decently in the 2017 election (Munzert et al. 2017; Stoetzer et al. 2019; see also http://zweitstimme.org). The model allows us to predict party-vote shares, coalition shares, the likelihood of a majority for certain coalitions, and many other relevant quantities of interest.

Our point of departure is a Bayesian forecasting approach that combines polls and fundamentals. This follows the tradition of synthetic forecasting models (Graefe 2017; Lewis-Beck and Dassonneville 2015), which combine the merits of fundamentals-based with poll-based models. Although dynamic versions of these models—which are updated by incorporating published polling data over time—have been applied especially in the US context (Erikson and Wlezien 2013; Heidemanns, Gelman, and Morris 2020; Linzer 2013; Silver 2020), their methodology does not easily transfer to multiparty settings. Forecasting the outcomes of multiparty elections poses particular challenges. We must predict simultaneously the support of multiple parties and, therefore, must account for the compositional nature of the data when modeling them.

This article applies our dynamic Bayesian forecasting model to predict the outcome of the 2021 German federal election. It systematically combines published pre-election public-opinion poll results with information from fundamentals-based forecasting models while also accounting for the dynamic evolution of party support in multiparty systems.

This article presents an early forecast of our model, calibrates it on the basis of historical data, and reports various quantities of interest, including the probabilities of a plurality of votes for a party, a majority of seats for certain coalitions in parliament, and the expected overall size of parliament due to the distribution of overhang (i.e., surplus) and compensatory seats. In its current form, the model generates forecasts for various points of interest beginning as early as 200 days before Election Day. Updated forecasts based on incoming polling information will be disseminated in the online edition of *Süddeutsche Zeitung*, a major German quality newspaper.

THE ZWEITSTIMME MODEL

As a synthetic forecasting model, the Zweitstimme model has two components (for a more detailed description, see Munzert et al. 2017 and Stoetzer et al. 2019). The mathematical details of this dynamic Bayesian measurement model are described in Stoetzer et al. (2019). For this symposium, we summarize the model's components.

The first component is a fundamentals-based model that forecasts each party's vote share. Existing fundamentals-based forecasting models use party-level predictors based on regularities in previous elections (Jérôme, Jérôme-Speziari, and Lewis-Beck 2013; Kayser and Leininger 2017; Norpoth and Gschwend 2010, 2017). Rather than linear or seemingly unrelated regression models, we use Dirichlet regressions with random effects to allow the effects of different predictors to vary over time while also accounting for the compositional nature of multiparty vote-share distribution. Our fundamentals-based model includes three covariates based on the idea that a few core factors should reasonably predict election outcomes: (1) long-term party attachment as a normal-vote baseline (i.e., previous election result); (2) short-term campaign dynamics (i.e., average vote intention in the polls 230 to 200 days before Election Day); and (3) an institutional feature that most heavily registers credit and blame regarding the incumbent government with support for the prime minister's party (i.e., dummy indicating the current chancellor's party). We fit the model using data on all postwar German federal elections. More details on the fundamentals-based model, including some of the results from the Dirichlet regression, are provided in the online appendix, section A.

doi:10.1017/S1049096521000913

The second component of our model is a dynamic Bayesian measurement model using published pre-election public-opinion polls by different polling companies. We model them as a multinomial process in which the count of respondents who intend to vote for a party is drawn from the total number of respondents with a certain probability. The probability to observe supporters for the different parties in the polls then is conceptualized as a combination of the latent support for the parties at a particular point in time and house effects of the different polling companies.[1] An additional feature of the Zweitstimme model is its dynamic process. We model the evolution of latent support for parties as a backward random walk, thereby extending existing two-party models used to forecast US presidential election outcomes (Linzer 2013) to multiparty systems. The backward random walk models the current latent support for each party as a linear combination of the latent support at the next point in time and random noise.[2]

Finally, both components are combined and jointly estimated in a dynamic Bayesian forecasting model. The main advantage of the backward-random-walk approach is that it allows integration of the forecasts from the fundamentals-based model. It isolates priors for the latent support of parties on Election Day that we set to the posterior predictive distribution from the fundamentals-based model.[3]

In its current form, our model generates forecasts for various quantities of interest as early as 200 days before Election Day. Evaluating our forecast of party vote shares of the 2017 Bundestag election, the forecast of the fundamentals-based model had a root mean squared error of 4.2 (with a lead time of 200 days). We further evaluated our combined model every four weeks beginning 148 days and two days before Election Day. Table 1 demonstrates that our combined model improves on the fundamentals-based forecast as Election Day nears.[4]

In addition to the previous German federal election in 2017 (Munzert et al. 2017; Stoetzer et al. 2019), we successfully applied it to elections in New Zealand (Stoetzer et al. 2019). The Zweitstimme model is implemented in Stan (Stan Development Team 2021) and the code is available on Harvard Dataverse (Neunhoeffer et al. 2018). The application to the 2021 election serves as an additional test for the Zweitstimme model and is particularly challenging considering the long lead time of the forecast and the recently increased volatility in the polls.

FORECAST FOR THE GERMAN FEDERAL ELECTIONS OF 2021

We use data from seven major German polling companies. Figure 1 lists our forecasts published as of June 17, 2021, 100 days before Election Day, along with the respective 5/6 (83%) credible intervals.[5] Accordingly, we predict that the Christian Democratic Union/Christian Social Union (CDU/CSU) will reach 31% (25%, 39%); the Social Democratic Party (SPD) 15% (11%, 19%); the Left Party 7% (5%, 9%); the Green Party 17% (12%, 23%); the Free Democratic Party (FDP) 12% (8%, 16%); the Alternative für Deutschland (AfD) 11% (8%, 14%); and Others 6% (5%, 8%).

Because the German federal election is still 100 days away (as of this writing), this article preregisters an early snapshot of our forecast. Figure 2 shows that early in the campaign, the fundamentals-based model substantially impacts the Election Day forecast, with the predictive distributions being centered around it. Currently, party support in the polls deviates from what we would expect from the fundamentals-based model; specifically, the CDU/CSU currently polls lower than what is predicted and the Green Party polls higher. Given that the forecast based on polls is still uncertain 100 days ahead, the information from the fundamentals-based model has a central role in anchoring the forecast.[6]

Nevertheless, we can provide some intuition on how our forecast—in particular, the associated uncertainty—will develop over the course of the campaign. Closer to Election Day, the polls become more informative and can pull the forecast away from the fundamentals-based model. At the same time, our predictions are more certain the closer we are to Election Day. For example, in the 2017 Bundestag election, the SPD forecast 148 days before Election Day was still heavily influenced by the fundamentals, expecting around 30% vote share for the SPD. However, it came with significant uncertainty, suggesting that support can decrease to as low as 20%. Indeed, during the campaign, the SPD percentage decreased; the fundamentals less strongly influenced the forecast; and the forecast became more precise, leading to a final forecast two days before

This article applies our dynamic Bayesian forecasting model to predict the outcome of the 2021 German federal election.

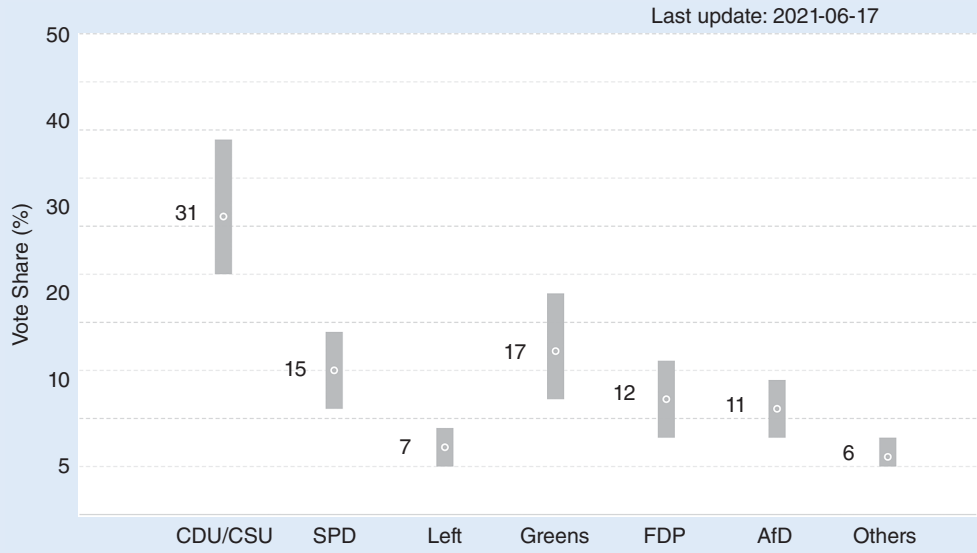
Table 1

Evaluation of our Forecast of the 2017 Bundestag Election

Lead time in days	Root mean squared error
2	1.9
8	2.1
36	3.3
64	3.4
92	3.3
116	3.4
148	4.3

Figure 1

Predictions for the 2021 Bundestag Election from the Zweitstimme Model



Election Day that was close to the 20.5% party vote share of the SPD in the 2017 Bundestag election (see figure A1 in the online appendix, section C.). We expect similar patterns in our forecasts to unfold during the 2021 campaign.

As of this writing, a distinct feature of the 2021 campaign seems to be that several ideologically plausible coalition options have a reasonable probability of receiving a majority of seats. This includes coalition governments without the CDU/CSU, which has led all governments (with changing coalition partners) since 2005. The probabilities for a majority of seats that our model currently (i.e., 100 days before Election Day) assigns to the following options are CDU/CSU–SPD 47%; CDU/CSU–Green Party 67%; CDU/CSU–FDP 20%; CDU/CSU–Green Party–FDP >99%; CDU/CSU–SPD–FDP >99%; Green Party–SPD–FDP 29%; and Green Party–SPD–Left Party 6%. Other options are either ideologically or statistically implausible.

DISTRICT-LEVEL FORECASTS AND THE SIZE OF THE BUNDESTAG

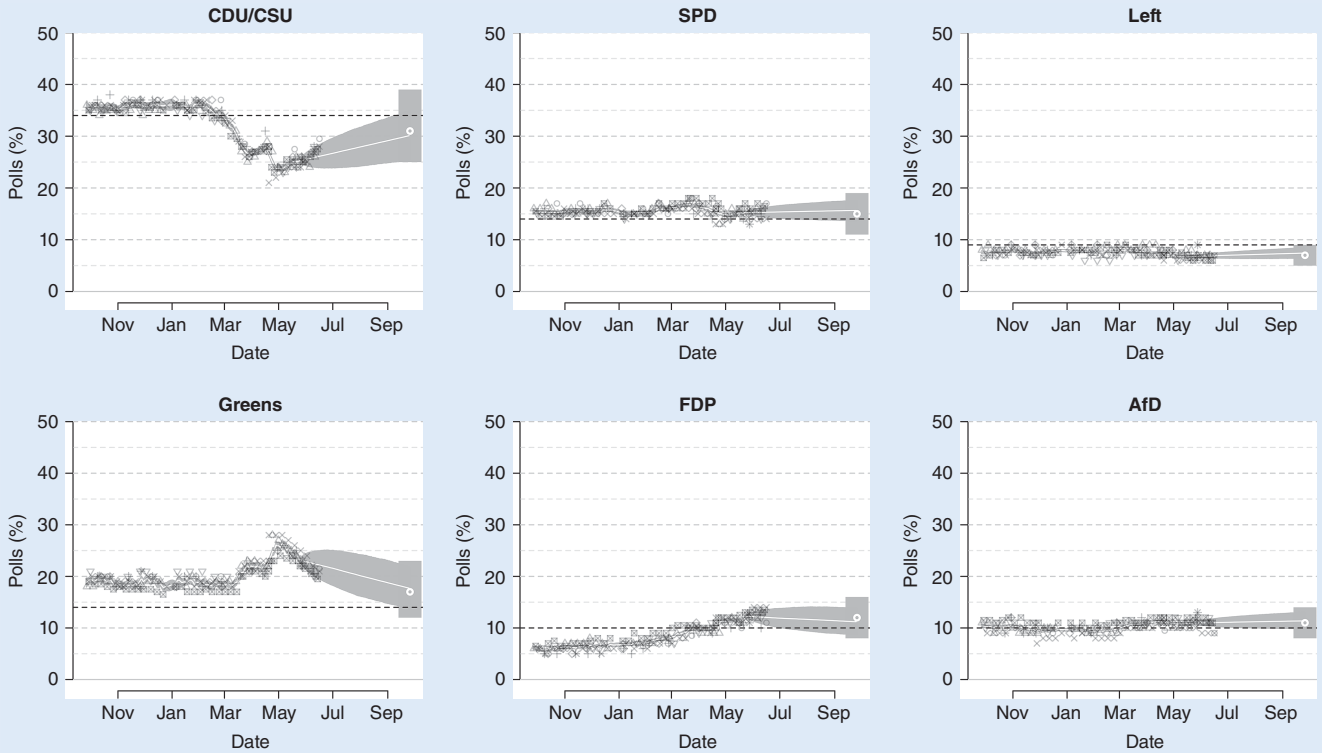
Germany has a mixed-member electoral system, in which voters can cast two votes. The first vote (*Erststimme*) is a nominal vote for district candidates; the second vote is a party vote (*Zweitstimme*). Because the latter vote is decisive for the distribution of seats in parliament, our Zweitstimme forecasting model focuses primarily on the party-vote distribution. However, important quantities of interest cannot be expressed when focusing only on the national vote share of parties, including the size of the Bundestag, the exact distribution of seats in parliament, and the candidates' chances to enter the Bundestag via the party list or a district vote. The online appendix, section B, describes how we combine our Zweitstimme forecast with an artificial neural network to generate district-level predictions using candidate- and district-level characteristics (Neunhoeffer et al. 2020). We illustrate the capability of our model to forecast district-level results by describing the forecast for one competitive district (i.e., 61 Potsdam) and describe the expected size of the Bundestag, which depends not only on the distribution of party votes nationally but also on the distribution of candidate votes across districts.[7] After conducting simulations for both—that is, predicted national party vote shares as well as candidate vote predictions in each electoral district—we use them according to electoral law to calculate the resulting size of parliament for each simulation.

Currently, we predict the Bundestag to have 814 seats with a 5/6 credible interval (726, 905), securing its reputation as one of the largest democratic legislative bodies (it will even outnumber the European Parliament's 705 members).

In Potsdam, the chancellor candidates Annalena Baerbock (Green Party) and Olaf Scholz (SPD) are competing as nominal candidates. Currently, the district-level candidates are not officially set. Until this occurs, our model assumes that all candidates for the 2021 election are those who already ran in

Figure 2

Vote Share Predictions Based on the Dynamic Bayesian Forecasting Model



Note: The symbols represent the party support reported in the respective polls (polling companies are indicated by different symbols). The black dashed line depicts the forecast of the fundamentals-based model. The shadowed area depicts the 5/6 credible intervals for the prediction of latent party support. The most recent update of the forecast was June 17, 2021.

2017. This is true for Baerbock but Scholz was not a candidate in this district in 2017. Considering our district-level prediction 100 days before Election Day, it is likely that neither chancellor candidate will win this district. The Green Party will have a 3% chance of winning, the SPD 7%. In contrast, the CDU candidate currently has an 88% chance to win this district. Figure A3 in the online appendix, sections C and D, provides an overview of forecasted district-level results for all electoral districts.

We also calculate the respective size of the Bundestag for each prediction.[8] Currently, we predict that the Bundestag will have 814 seats with a 5/6 credible interval (726, 905), securing its reputation as one of the largest democratic legislative bodies (it will even outnumber the European Parliament's 705 members). Compared to the current Bundestag with 709 seats, the next Bundestag likely will increase in size and have more than 709 seats, with a probability of 94%. As a point of reference, the legal target size for the Bundestag is 598 seats. The mixed-member electoral system makes additional seats necessary because the overall seat-share distribution should closely match the distribution of party votes. If a party were to win more electoral districts than seats it would be entitled to by the party vote (based on 598 seats), the Bundestag would need additional seats to appropriately map the distribution of party votes. Our forecast addresses current debates on election-system reform that aim to solve the issue of an increase in the size of the Bundestag. The coalition CDU/CSU—SPD government proposed such a reform with alterations for the 2021 German federal elections and more extensive adjustments for subsequent elections. However, our prediction shows that this reform will not successfully prevent the increase in the size of the parliament: according to the coalition's reform, the next Bundestag is expected to be smaller by only nine (6, 35) seats compared to when the former electoral law is applied. Considering this in addition to the likely chance of the upcoming Bundestag being larger than the current one, the effectiveness of the proposed reform, in fact, is questionable.

DISCUSSION

This article presents the Zweitstimme model to forecast the German federal election of 2021. To summarize, we expect a strengthening in Green Party support, which opens up new coalition options. Adding district-level forecasts further allows predicting additional points of interest, such as close district races and the size of the Bundestag. Evaluating all politically plausible[9] and arithmetically possible coalition governments that could form after the election, our model currently predicts an 88% chance that Armin Laschet (CDU/CSU) will lead the next German government, Annalena Baerbock (Green Party) has an 8% chance to succeed Angela Merkel, and Olaf Scholz (SPD) has a 3% chance.

Forecasting models are never carved in stone. An interesting addition to our model could be the prediction of coalition options—not merely hypothetical coalition-government-seat majorities. The integration of models for coalition bargaining behavior could allow us to obtain a true forecast of the coalition government and, in turn, the likely chancellor. Furthermore, the dynamics of our polls model currently are based on backward random walks. Including more elaborate dynamics might prove effective in obtaining more information from the polls weeks ahead of Election Day.

Evaluating all politically plausible and arithmetically possible coalition governments that could form after the election, our model currently predicts an 88% chance that Armin Laschet (CDU/CSU) will lead the next German government.

DATA AVAILABILITY STATEMENT

Research documentation and data that support the findings of this study are openly available at the *PS: Political Science & Politics* Dataverse: https://doi.org/10.7910/DVN/EDTKNW.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit https://doi.org/10.7910/DVN/EDTKNW. ■

NOTES

1. House effects consider that results reported by some of the polling companies might be systematically biased in favor or disfavor of the different parties. The model relies on a log-ratio transformation to ensure that the linear combination of the latent support and house effects maps onto the probability interval for the multinomial process.
2. In the multiparty setting, we model this random noise as stemming from a multivariate normal distribution and the respective variance–covariance matrix represents variability and covariance of the evolution of the latent support over time.
3. The full posterior distribution requires specifications of the prior distributions. We use the same prior distributions as in Stoetzer et al. (2019).
4. For more information on the evolution of our forecast in the context of the 2017 Bundestag election, see figure A2 in the supplementary materials.
5. Replication data and code for all of the results in this article are available in Neunhoeffer et al. (2021).
6. Figure 2 further highlights that the forecast intervals are wider than the intervals from the dynamic trends on Election Day. The evaluation based on past elections revealed that the forecasts on Election Day from the dynamic process alone were overconfident. We use the same correction as in Stoetzer et al. 2019, adding additional error variation to the forecast of each party vote share on Election Day.
7. The district-level prediction model described in Neunhoeffer et al. (2020) requires information on all candidates in each of the 299 electoral districts. Because this information is not yet available for the 2021 election (as of June 17, 2021), we must make additional simplifying assumptions. Most important for the current district-level prediction, we assume that the same candidates who ran in 2017 will run again in 2021 (because this is the best data available). This allows a first prediction for the number of seats in the 2021 Bundestag. As soon as the data on the 2021 candidates are available, we will publish more detailed district-level predictions in the online edition of *Süddeutsche Zeitung*.
8. Given simulated district winners and simulated party vote shares, we can directly apply the electoral rules spelled out by the Bundeswahlleiter to each simulated result to obtain a seat distribution. The demo calculation is available

at www.bundeswahlleiter.de/dam/jcr/05c1185a-173f-4bab-80d6-51027c94b1bc/bwg2021_mustersitzberechnung_ergebnis2017.pdf (accessed May 28, 2021).

9. Politically plausible and arithmetically possible (at least in some simulations) are the following coalition options: a CDU/CSU majority, a CDU/CSU–SPD coalition, a CDU/CSU–Green Party coalition, a CDU/CSU–FDP coalition, a Green Party–SPD coalition, a CDU/CSU–Green Party–FDP coalition, a CDU/CSU–SPD–FDP coalition, a Green Party–SPD–FDP coalition, and a Green Party–SPD–Left Party coalition. The order of the parties follows our current forecast but may be different in some simulations.

REFERENCES

Erikson, Robert S., and Christopher Wlezien. 2013. "Forecasting with Leading Economic Indicators and the Polls 2012." *PS: Political Science & Politics* 46 (1): 38–39.

Graefe, Andreas. 2017. "The PollyVote's Long-Term Forecast for the 2017 German Federal Election." *PS: Political Science & Politics* 50 (3): 693–95.

Heidemanns, Merlin, Andrew Gelman, and G. Elliott Morris. 2020. "An Updated Dynamic Bayesian Forecasting Model for the US Presidential Election." *Harvard Data Science Review* 2 (4). https://doi.org/10.1162/99608f92.fc62f1e1.

Jérôme, Bruno, Véronique Jérôme-Speziari, and Michael S. Lewis-Beck. 2013. "A Political-Economy Forecast for the 2013 German Elections: Who to Rule with Angela Merkel?" *PS: Political Science & Politics* 46 (3): 479–80.

Kayser, Mark A., and Arndt Leininger. 2017. "A *Länder*-Based Forecast of the 2017 German Bundestag Election." *PS: Political Science & Politics* 50 (3): 689–92.

Lewis-Beck, Michael S., and Ruth Dassonneville. 2015. "Forecasting Elections in Europe: Synthetic Models." *Research & Politics* 2 (1): 1–11.

Linzer, Drew A. 2013. "Dynamic Bayesian Forecasting of Presidential Elections in the States." *Journal of the American Statistical Association* 108 (501): 124–34.

Munzert, Simon, Lukas F. Stoetzer, Thomas Gschwend, Marcel Neunhoeffer, and Sebastian Sternberg. 2017. "Zweitstimme.org. Ein Strukturell-Dynamisches Vorhersagemodell Für Bundestagswahlen." *Politische Vierteljahresschrift* 58 (3): 418–41.

Neunhoeffer, Marcel, Thomas Gschwend, Klara Müller, Simon Munzert, and Lukas F. Stoetzer. 2021. "Replication Data for: The Zweitstimme Model: A Dynamic Forecast of the 2021 German Federal Election." Harvard Dataverse. https://doi.org/10.7910/DVN/EDTKNW.

Neunhoeffer, Marcel, Thomas Gschwend, Simon Munzert, and Lukas F. Stoetzer. 2020. "Ein Ansatz Zur Vorhersage der Erststimmenanteile bei Bundestagswahlen." *Politische Vierteljahresschrift* 61 (1): 111–30.

Neunhoeffer, Marcel, Lukas F. Stoetzer, Thomas Gschwend, Simon Munzert, and Sebastian Sternberg. 2018. "Replication Data for: Forecasting Elections in Multi-Party Systems: A Bayesian Approach Combining Polls and Fundamentals." Harvard Dataverse. https://doi.org/10.7910/DVN/MLYNXo.

Norpoth, Helmut, and Thomas Gschwend. 2010. "The Chancellor Model: Forecasting German Elections." *International Journal of Forecasting* 26 (1): 42–53.

Norpoth, Helmut, and Thomas Gschwend. 2017. "Chancellor Model Predicts a Change of the Guards." *PS: Political Science & Politics* 50 (3): 686–88.

Silver, Nate. 2020. "FiveThirtyEight 2020 US Presidential Election Forecast." https://projects.fivethirtyeight.com/2020-election-forecast.

Stan Development Team. 2021. "Stan Modeling Language User's Guide and Reference Manual, 2.26." https://mc-stan.org.

Stoetzer, Lukas F., Marcel Neunhoeffer, Thomas Gschwend, Simon Munzert, and Sebastian Sternberg. 2019. "Forecasting Elections in Multiparty Systems: A Bayesian Approach Combining Polls and Fundamentals." *Political Analysis* 27 (2): 255–62.