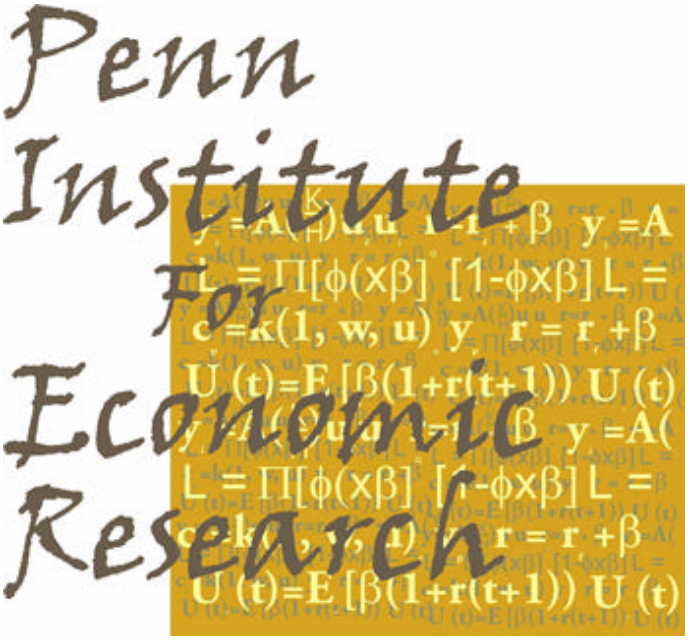


Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 02-046

"Weather Forecasting for Weather Derivatives"

by

Sean D. Campbell and Francis X. Diebold

http://ssrn.com/abstract_id=359663

Weather Forecasting for Weather Derivatives

Sean D. Campbell
Brown University

Francis X. Diebold
University of Pennsylvania
and NBER

December 2000
This revision/print: December 4, 2002

Abstract: We take a nonstructural time-series approach to modeling and forecasting daily average temperature in ten U.S. cities, and we inquire systematically as to whether it may prove useful from the vantage point of participants in the weather derivatives market. The answer is, perhaps surprisingly, yes. Time series modeling reveals both strong conditional mean dynamics and conditional variance dynamics in daily average temperature, and it reveals sharp differences between the distribution of temperature and the distribution of temperature surprises. Most importantly, it adapts readily to produce the long-horizon forecasts of relevance in weather derivatives contexts. We produce and evaluate both point and distributional forecasts of average temperature, with some success. We conclude that additional inquiry into nonstructural weather forecasting methods, as relevant for weather derivatives, will likely prove useful.

Key Words: Risk management; hedging; insurance; seasonality; average temperature; financial derivatives; density forecasting

JEL Codes: G0, C1

Acknowledgments: For financial support we thank the National Science Foundation, the Wharton Financial Institutions Center, and the Wharton Risk Management and Decision Process Center. For helpful comments we thank Marshall Blume, Larry Brown, Jeff Considine, John Dutton, René Garcia, Stephen Jewson, Vince Kaminski, Paul Kleindorfer, Howard Kunreuther, Yu Li, Bob Livezey, Cliff Mass, Don McIsaac, Nour Meddahi, David Pozo, Matt Pritsker, S.T. Rao, Claudio Riberio, Til Schuermann and Yihong Xia. We are also grateful for comments by participants at the American Meteorological Society's Policy Forum on Weather, Climate and Energy. None of those thanked, of course, are responsible in any way for the outcome.

Address correspondence to:
F.X. Diebold
Department of Economics
University of Pennsylvania
3718 Locust Walk, Philadelphia, PA 19104-6297
fdiebold@sas.upenn.edu

1. Introduction

Weather derivatives are a fascinating new type of Arrow-Debreu security, making pre-specified payouts if pre-specified weather events occur. The market has grown rapidly. In 1997, the market for weather derivatives was nonexistent. In 1998 the market was estimated at $500 million, but it was still illiquid, with large spreads and limited secondary market activity. More recently the market has grown to more than $5 billion, with better liquidity. Outlets like *Energy and Power Risk Management* and the *Weather Risk* (e.g., 1998, 2000) supplements to *Risk Magazine* have chronicled the development.[1]

Weather derivative instruments include weather swaps, vanilla options, option collars, and exotic (e.g., path-dependent) options. The underlying include heating degree days, cooling degree days, growing degree days, average temperature, maximum temperature, minimum temperature, precipitation (rainfall, snowfall), humidity and sunshine, among others – even the National Weather Service's temperature forecast for the coming week.[2]

A number of interesting considerations make weather derivatives different from "standard" derivatives. First, the underlying (weather) is not traded in a spot market. Second, unlike financial derivatives, which are useful for price hedging but not quantity hedging, weather derivatives are useful for quantity hedging but not necessarily for price hedging (although the two are obviously related). That is, weather derivative products provide protection against weather-related changes in quantities, complementing extensive commodity price risk management tools already available through futures. Third, although liquidity in weather derivative markets has improved, it will likely never be as good as in traditional price-hedging markets, because weather is by its nature a location-specific and non-standardized commodity, unlike, say, a specific grade of crude oil.

Weather derivatives are also different from insurance. First, there is no need to file a claim or prove damages. Second, there is little moral hazard, although there is some, as when someone with a long precipitation position attempts to seed the clouds. (Don't laugh – it has happened!) Third, unlike insurance, weather derivatives allow one to hedge against comparatively good weather in other locations, which may be *bad* for local business (e.g., a bumper crop of California oranges may lower the prices received by Florida growers).

Weather forecasting is crucial to both the demand and supply sides of the weather derivatives market. Consider first the demand side, consisting of obvious players such as energy companies, utilities and insurance companies, and less obvious players such as ski resorts, grain millers, cities facing snow-

---

[1] For an interesting overview, see Geman (1998).

[2] Aquila Energy's "Guaranteed Forecast" weather hedging product structures put and call options around the National Weather Service's temperature forecast.

removal costs, consumers who want fixed heating and air conditioning bills, firms seeking to avoid financial writedowns due to weather-driven poor performance, investors seeking an approximately "zero beta" asset class, etc.[3] The mere fact that such agents face weather fluctuations, however, does not ensure a large hedging demand, because even very large weather fluctuations would create little weather risk if they were highly predictable. Weather risk, then, is about the *un*predictable component of weather fluctuations – "weather surprises," or "weather noise." To assess the potential for hedging against weather surprises, and to formulate the appropriate hedging strategies, one needs to determine how much weather noise exists for weather derivatives to eliminate, and that requires a weather model. What does weather noise look like over space and time? What are its conditional and unconditional distributions? Answering such questions requires weather modeling and forecasting, the topic of this paper.

Now consider the supply side – sellers of weather derivatives who want to price them, statistical arbitrageurs who want to exploit situations of apparent mispricing, etc. How should weather derivatives be priced? It seems clear that standard approaches to arbitrage-free pricing (e.g., Black-Scholes) are inapplicable in weather derivative contexts. In particular, there is in general no way to construct a portfolio of financial assets that replicates the payoff of a weather derivative (although in some situations one might be able to construct a crude approximation to a replicating portfolio if futures are traded). Hence the only way to price options reliably is by using forecasts of the underlying weather variable, in conjunction with a utility function. As Davis (2001, p. 305) notes:[4]

> Since there is no liquid market in these contracts, Black-Scholes style pricing is inappropriate. Valuation is generally done on an "expected discounted value" basis, discounting at the riskless rate but under the physical measure, which throws all the weight back onto the problem of weather prediction.

This again brings up the crucial issue of how to construct good weather forecasts, potentially at horizons much longer than those commonly emphasized by meteorologists. Hence the supply-side questions, as with the demand-side questions, are intimately related to weather modeling and forecasting.

Curiously, however, it seems that little thought has been given to the crucial question of how best to approach the weather modeling and forecasting that underlies weather derivative demand and supply. The extant weather forecasting literature (see, for example, the overview in Tribia, 1997) has a structural "atmospheric science" feel, and although such an approach may be best for forecasting six hours ahead, it

---

[3] Effectively, any firm exposed to weather risk either on the output (revenue) side or the input (cost) side is a candidate for productive use of weather derivatives.

[4] See also Cao and Wei (2001).

is not at all obvious that it is best for the longer horizons relevant for weather derivatives, such as six weeks or six months. In particular, successful forecasting does not necessarily require a structural model, and in the last thirty years statisticians and econometricians have made great strides in the nonstructural modeling and forecasting of time series trend, seasonal, cyclical and noise components.[5] The time series approach has proved useful not only in providing accurate "best guesses" (point forecasting) but also in providing detailed quantification of associated forecast uncertainty (distributional forecasting). In this paper, then, motivated by considerations related to the weather derivatives market, we take a time-series approach to weather modeling and forecasting, systematically asking whether the nonstructural approach proves useful.[6] We provide insight into both conditional mean dynamics and conditional variance dynamics of daily average temperature; strong conditional variance dynamics are a central part of the story. We also highlight the differences between the distributions of weather and the distributions of weather surprises. Finally, we evaluate the performance of our model's distributional forecasts at the long horizons of relevance in weather derivatives contexts. The results are mixed but encouraging, and they point toward directions that may yield future forecasting improvements.

We proceed as follows. In Section 2 we discuss our data, and our decision to focus on modeling and forecasting daily average temperature. In section 3 we report the results of time-series modeling of daily average temperature, and in section 4 we report the results of out-of-sample point forecasting and distributional forecasting exercises. In section 5 we offer concluding remarks and sketch directions for future research.

2. Data and Modeling Choices

Here we discuss the collection and cleaning of our weather data, the subset of the data that we model, and the reasons for our choice.

Data

Our dataset contains daily temperature observations measured in degrees Fahrenheit for each of the ten measurement stations listed in Table 1 for 1/1/60 through 11/05/01, resulting in 15,285

[5] See Diebold (2001).

[6] We are not the first to adopt a time-series approach, although the literature is sparse. The analyses of Harvey (1989), Hyndman and Grunwald (2000), Milionis and Davies (1994), Visser and Molenaar (1995), Jones (1996), and Pozo et al. (1998) suggest its value, for example, but they do not address the intra-year temperature forecasting relevant to our concerns. Seater (1993) studies long-run temperature trend, but little else. Cao and Wei (2001) and Torro, Meneu and Valor (2001) – each of which was written independently of the present paper – consider time-series models of average temperature, but their models are more restrictive and their analyses more limited than ours.

observations per measuring station. We chose the stations to match those on which weather-related futures and options are traded at the Chicago Mercantile Exchange (CME). We obtained the data from Earth Satellite (EarthSat) corporation; they are precisely those used to settle temperature-related weather derivative products traded on the CME. The primary underlying data source is the National Climactic Data Center (NCDC), a division of the National Oceanographic and Atmospheric Administration. Each of the measuring stations listed in Table 1 supplies its data to the NCDC, and those data are in turn collected by EarthSat. The dataset consists of observations on the temperature-related variables detailed in Table 2: daily maximum (*TMAX*), minimum (*TMIN*), and average (*T*) temperatures, together with heating degree days (*HDD*) and cooling degree days (*CDD*), the formulae for which are given in the table.

The dataset contains occasional missing observations due to the failure of measuring stations to report to the NCDC. Such instances are rare and are detailed in Table 3; they occur just twelve times in our sample and are attributable to factors such as human error or mechanical failure of the measurement equipment. When missing values are encountered, we (and the CME) use fitted values constructed by EarthSat as follows. First, EarthSat identifies the geographically closest NCDC measuring station (reference station) for each of the ten cities contained in the data set, as shown Table 3. Second, EarthSat calculates for each city the thirty-year daily average difference of the missing variable (*TMAX* or *TMIN*) between the measuring station and its reference station.[7] In this calculation, each day in the year is taken as distinct; hence the thirty-year average is based on thirty observations. The average differences are also shown in Table 3, as are the sample standard deviations of the differences. Finally, EarthSat adds to the reference station measurement the thirty-year average difference.

Modeling Daily Average Temperature

We are interested in daily average temperature, which is widely reported and followed. Moreover, the heating degree days (*HDD*s) and cooling degree days (*CDD*s) on which weather derivatives are commonly written are simple transformations of average temperature, as shown in Table 2.

Perhaps surprisingly, a key issue in modeling and forecasting daily average temperature is precisely what to model. From one point of view, it seems unnecessary and potentially undesirable to model separately weather variables such as daily average, maximum and minimum temperature, *HDD*s, *CDD*s, etc., given that one can simply model the underlying high-frequency temperature, from which all

[7] Only missing values for *TMAX* and *TMIN* need be filled, as all of the other weather variables are derived from them.

else is derived. But even if we had access to the relevant high-frequency data (which we do not), once model misspecification is acknowledged, it often proves attractive to work under the relevant loss function, modeling and forecasting the object of interest directly. The attractiveness stems both from the fact that the requisite modeling is much simpler (e.g., there is no need to model complicated intra-day calendar effects) and from the fact that the resulting forecasts are tailored to the object of interest and may therefore be more accurate for the object of interest.[8]

This same logic of forecasting under the relevant loss function suggests that, although the daily maximum and minimum temperatures (to which we *do* have access) are the fundamental daily series from which all others are derived, as revealed in Table 2, it does not necessarily follow that we should model and forecast *TMAX* and *TMIN* and then derive the implied forecast for daily average temperature. Again, direct modeling of daily average temperature is certainly simpler and may produce more accurate forecasts of daily average temperature.

The upshot: we model and forecast daily average temperature.

3. Time-Series Weather Modeling

Before proceeding to detailed modeling and forecasting results, it is useful to get an overall feel for the daily average temperature data. In Figure 1 we plot the average temperature series for the last five years of the sample.[9] The time series plots reveal strong and unsurprising seasonality in average temperature: in each city, the daily average temperature moves repeatedly and regularly through periods of high temperature (summer) and low temperature (winter). Importantly, however, the seasonal fluctuations differ noticeably across cities both in terms of amplitude and detail of pattern.

In Figure 2 we show how the seasonality in daily average temperature manifests itself in the unconditional temperature distributions. Most cities' distributions are bimodal, with peaks characterized by cool and warm temperatures. Also, with the exception of Las Vegas, each distribution is negatively skewed.[10] The daily average temperature distributions for the Midwestern cities of Chicago and Des Moines have the greatest dispersion, while the western cities of Portland and Tucson have the least

---

[8] For a recent treatment of forecasting under specialized loss functions, see Christoffersen and Diebold (1996, 1997).

[9] We could of course have plotted the average temperature series over the entire sample, but doing so compresses too much data into too little space, resulting in a less informative graph.

[10] The distributions are in line with von Storch and Zwiers (1999), who note that although daily average temperature often appears Gaussian if studied over sufficiently long times in the troposphere, daily average surface temperatures may have different distributions. See also Neese (1994), who documents skewness and bimodality in daily maximum temperatures.

dispersion.

The discussion thus far suggests that a seasonal component will be important in any time-series model fit to daily average temperature, as average temperature displays pronounced seasonal variation, with both the amplitude and precise seasonal patterns differing noticeably across cities. We use a low-ordered Fourier series as opposed to daily dummies to model this seasonality, the benefits of which are two-fold. First, use of a low-ordered Fourier series produces a smooth seasonal pattern, which accords with the basic intuition that the progression through different seasons is gradual rather than discontinuous. Second, the Fourier approximation produces a huge reduction in the number of parameters to be estimated, which significantly reduces computing time and enhances numerical stability. Such considerations are of relevance given the rather large size of our dataset (roughly 15,000 daily observations for each of ten cities).

One naturally suspects that non-seasonal factors may also be operative in the dynamics of daily average temperature, despite the fact that their relevance is not obvious from the simple time series plots examined thus far, which are dominated by seasonality. One obvious such factor is trend, which may be relevant but is likely minor, given the short forty-year span of our data. We therefore simply allow for a deterministic linear trend. Another such factor is cycle, by which we mean any sort of persistent (but covariance stationary) dynamics apart from seasonality and trend. We capture cyclical dynamics using autoregressive lags, which facilitates fast and numerically-stable computation of parameter estimates.[11]

Assembling the various pieces, we estimate the following model for average temperature in each of our ten cities between 1/1/1960 and 11/05/2001:

$$T_t = Trend_t + Seasonal_t + \sum_{l=1}^{L} \rho_{t-l} T_{t-l} + \sigma \varepsilon_t, \tag{1}$$

where

$$Trend_t = \beta_0 + \beta_1 t \tag{1a}$$

$$Seasonal_t = \sum_{p=1}^{P} \left(\delta_{c,p} \cos\left(2\pi p \frac{d(t)}{365}\right) + \delta_{s,p} \sin\left(2\pi p \frac{d(t)}{365}\right) \right) \tag{1b}$$

[11] ARMA models would in principle enable greater parsimony, but only at the expense of requiring numerical optimization for parameter estimation.

$$\varepsilon_t \overset{iid}{\sim} (0, 1), \tag{1c}$$

and where *d(t)* is a repeating step function that cycles through 1, ..., 365 (i.e., each day of the year assumes one value between 1 and 365).[12] We estimate the model by ordinary least squares, regressing average temperature on constant, trend, Fourier and lagged average temperature terms, using twenty-five autoregressive lags (L=25) and three Fourier sine and cosine terms (P=3).[13]

Now let us discuss the estimation results. First, the seasonality is of course statistically significant and crucially important. To assess the Fourier fit, we plot in Figure 3 the estimated Fourier seasonal pattern against an estimated dummy variable seasonal pattern, corresponding to $Seasonal_t = \sum_{i=1}^{365} \delta_i D_{it}$, where $D_{it}=1$ on the i *-th* day of the year, and 0 otherwise. In general, the fit is spectacular; the noise in the dummy variable seasonal pattern is greatly reduced with almost no deterioration of fit.

Second, and perhaps surprisingly, most cities display a large and statistically significant trend in daily average temperature. In most cases, the trend is much larger than the increase in average global temperature over the same period. For example, the results indicate that the daily average temperature in Atlanta has increased by three degrees in the last forty years. Such large trend increases are likely a consequence of development and air pollution that increased urban temperatures in general, and urban airport temperatures in particular, where most of the U.S. recording stations are located, a phenomenon often dubbed the "heat island effect."[14]

Daily average temperature also displays strong cyclical persistence. The estimated autoregressive coefficients display an interesting pattern, common across all ten cities, regardless of location. The coefficient on the first lag is typically large and significant, around 0.85, but coefficients on subsequent

---

[12] We removed February 29 from each leap year in our sample to maintain 365 days per year. Alternatively, we could set a "year" equal to 365.25 days.

[13] For economical presentation we simply report results with L=25 and P=3 for all cities. We also used the Akaike and Schwarz information criteria to guide city-specific model selection, with little change in the forecasting results subsequently reported. For all cities, both criteria indicate that setting L=25 and P=3 is more than adequate. Maintaining the rather large value of L=25 costs little given the large number of available degrees of freedom, and it helps to capture long-memory dynamics, if present, and as indicated in results such as Bloomfield (1992).

[14] The fitted trend could also be approximating a very low-frequency cycle, although that seems unlikely given the large number of autoregressive lags included in the models.

lags become significantly negative before decaying. The coefficient on the second autoregressive lag, for example, is typically about -0.25. All roots of the associated autoregressive lag operator polynomials are less than one in modulus; the dominant root is always real and typically around 0.85, and the vast majority of remaining roots are complex, as shown in Figure 4.

In Figure 5 we show fitted values and residuals over the last five years of the estimation sample. The fit is typically very good, with R^2 above 90%. Figure 5 also provides a first glimpse of an important phenomenon: pronounced and persistent time-series variation in the variance of the residuals. Weather risk, as measured by its innovation variance, is often time-varying: cities such as Atlanta, Dallas and Des Moines show clear signs of seasonal heteroskedasticity as the range of the residuals widens and narrows over the course of a year.[15]

In Figure 6 we show residual histograms for each of the models. Four features stand out. First, average temperature residuals are much less variable than average temperature itself; that is, weather surprises are much less variable than the weather itself.[16] Second, the spreads of the residual distributions vary noticeably across cities; the standard deviations range from a low of 3.5 degrees for Tucson to a high of 6.2 degrees for Des Moines, indicating that weather isk is much greater in some cities than in others. Third, all of the distributions are uni- as opposed to bi-modal, in contrast to the unconditional distributions of daily average temperature examined earlier, due to the model's success in capturing seasonal highs and lows. Fourth, all of the distributions have moderate negative skewness and moderate excess kurtosis.

In Figures 7 and 8 we display the correlograms of the residuals and squared residuals, taken to a maximum displacement of 800 days. The residual autocorrelations are negligible and appear consistent with white noise, indicating that we have modeled linear dependence adequately. Figure 8, however, reveals drastic misspecification of the model (1), related to *non*linear dependence. With the exceptions of Las Vegas, Portland and Tucson (all west coast cities), the correlograms of the squared residuals show strong persistence, which highlights the need to incorporate conditional heteroskedasticity in the model.[17]

---

[15] It seems that such seasonal heteroskedasticity was first noted, informally, in an economic context by Roll (1984). In this paper we progress by developing quantitative models and forecasts that explicitly incorporate the heteroskedasticity.

[16] Typical residual standard deviations are only one third or so of the average temperature standard deviations.

[17] Conditional heteroskedasticity in the residuals of our (linear) model could also be indicative of neglected conditional-mean nonlinearities, but we have not yet explored that possibility.

In light of the clear residual heteroskedasticity, we move to the model,[18]

$$T_t = Trend_t + Seasonal_t + \sum_{l=1}^{L} \rho_{t-l} T_{t-l} + \sigma_t \varepsilon_t, \quad (2)$$

where

$$Trend_t = \beta_0 + \beta_1 t \quad (2a)$$

$$Seasonal_t = \sum_{p=1}^{P} \left(\delta_{c,p} \cos\left(2\pi p \frac{d(t)}{365}\right) + \delta_{s,p} \sin\left(2\pi p \frac{d(t)}{365}\right) \right) \quad (2b)$$

$$\sigma_t^2 = \sum_{q=1}^{Q} \left(\gamma_{c,q} \cos\left(2\pi q \frac{d(t)}{365}\right) + \gamma_{s,q} \sin\left(2\pi q \frac{d(t)}{365}\right) \right) + \sum_{r=1}^{R} \alpha_r \varepsilon_{t-r}^2 \quad (2c)$$

$$\varepsilon_t \overset{iid}{\sim} (0, 1), \quad (2d)$$

and where, as before, *d(t)* is a repeating step function that cycles through 1, ..., 365 (i.e., each day of the year assumes one value between 1 and 365), and we set L=25, P=3, Q=2 and R=1.[19]

Model (2) is of course identical to model (1), with the addition of the conditional variance equation (2c), which allows for two types of volatility dynamics relevant in time-series contexts. The first is volatility seasonality, which was visually apparent in many of the residual plots in Figure 5. Equation (2c) approximates seasonality in the conditional variance in precisely the same way as equation (2b) approximates seasonality in the conditional mean, via a Fourier series. The second is autoregressive effects in the conditional variance movements, which often arise naturally in time-series contexts, in which shocks to the conditional variance may have effects that persist for several periods, precisely as in the seminal work of Engle (1982).

We estimate the models by Engle's (1982) asymptotically efficient two-step approach, as follows. First, we estimate (2) by ordinary least squares, regressing average temperature on constant, trend, Fourier and lagged average temperature terms. Second, we estimate the variance equation (2c) by regressing the

---

[18] Figures 3-8 were actually produced using the results from model (2), in order to enhance the efficiency of parameter estimation once the above-discussed diagnostics made clear that the residuals were conditionally heteroskedastic.

[19] Just as with L and P, we also used a variety of information-theoretic strategies for selecting Q and R, but in the end chose simply to report results for Q=2 and R=1, which proved adequate for each city.

squared residuals from equation (2) on constant, Fourier and lagged squared residual terms, and we use the square root of the inverse fitted values, $\hat{\sigma}_t^{-1}$, as weights in a weighted least squares re-estimation of equation (2).

In Figure 9 we plot the estimated residual conditional standard deviation from 1996 through 2001. The basic pattern is one of strong seasonal volatility variation, with short-lived upward-brushing ARCH effects.[20] For each city, seasonal volatility appears highest during the winter months. Among other things, this indicates that correct pricing of weather derivatives may in general be crucially dependent on the season covered by the contract. Some cities display a great deal of seasonal volatility variation – the conditional standard deviations in Atlanta, Cincinnati and Dallas, for example, are roughly triple in winter – whereas weather surprise volatility in other cities such as Las Vegas, Portland and Tucson appears considerably more stable across seasons. Notice, moreover, that the amplitude of seasonal volatility fluctuations increases with its average level. That is, cities with high average volatility such as Atlanta and Dallas display the most pronounced seasonal volatility fluctuations, while cities with low average volatility such as Tucson and Portland also display smaller seasonal volatility fluctuations.

In Figure 10 we show histograms of standardized residuals, $(T_t - \hat{T}_t)/\hat{\sigma}_t$. The histograms reveal that, as expected, each standardized residual distribution still displays negative skewness. Modeling the conditional heteroskedasticity does, however, reduce (but not eliminate) residual excess kurtosis.[21]

In Figures 11 and 12 we show the correlograms of standardized and squared standardized residuals. The correlograms of standardized residuals are qualitatively similar to those in Figure 7 – there was no evidence of serial correlation in the residuals from (1), and hence there is no evidence of serial correlation in the standardized residuals from (2). The correlograms of the *squared* standardized residuals from model (2), however, show radical improvement relative to those from model (1); there is no significant deviation from white noise behavior, indicating that the fitted model (2) is adequate.

4. Time-Series Weather Forecasting

Armed with a hopefully adequate nonstructural time-series model for daily average temperature, we now proceed to examine its performance in out-of-sample weather forecasting. We assess our

---

[20] In sharp contrast to the standard situation in financial econometrics (see, for example, Andersen, Bollerslev and Diebold, 2003), it appears that low-ordered pure ARCH models are adequate for capturing weather volatility dynamics. We allowed for "GARCH lags" in our conditional variance specification, but none are necessary, so long as the volatility seasonality is modeled adequately. If the highly persistent volatility seasonality is not modeled adequately, however, then highly persistent (and spurious) GARCH effects may appear in volatility.

[21] It is not obvious, however, how to interpret the kurtosis of a skewed distribution.

performance relative to a number of competitors, including a leading structural meteorological forecast. One naturally suspects that the much larger information set on which the meteorological forecast is based will result in superior short-horizon point forecasting performance, but even if so, of great interest is the question of how quickly and with what pattern the superiority of the meteorological forecast deteriorates with forecast horizon. And more generally, of course, the options pricing considerations relevant in weather derivatives contexts suggest the centrality of distributional, as opposed merely to point, forecasts. We shall examine both.

Point Forecasting

We assess the short- to medium-term accuracy of daily average temperature forecasts based on our seasonal+trend+cycle model. In what follows, we refer to those forecasts as "autoregressive forecasts," for obvious reasons. We evaluate the autoregressive forecasts relative to three benchmark competitors, which range from rather naive to very sophisticated.

The first benchmark forecast is a no-change forecast. The no-change forecast, often called the "persistence forecast" in the climatological literature, is the minimum mean squared error forecast at all horizons if daily average temperature follows a random walk.

The second benchmark forecast is from a more sophisticated two-component (seasonal+trend) model. It captures (daily) seasonal effects daily effects via day-of-year dummy variables, in keeping with the common climatological use of daily averages as benchmarks. It captures trend via a simple linear deterministic function of time, which appears adequate for modeling the earlier-discussed long-term gradual warming at recording stations. We refer to this forecast as the "climatological forecast."

The third benchmark forecast, unlike benchmarks one and two, is not at all naive; on the contrary, it is a highly sophisticated forecast produced in real time by EarthSat. To produce their forecast, EarthSat meteorologists pool their expert judgement with model-based numerical weather prediction (NWP) forecasts from the National Weather Service, as well as forecasts from European, Canadian, and U.S. Navy weather services. This blending of judgement with models is typical of best-practice modern weather forecasting.

We were able to purchase approximately two years of forecasts from EarthSat. The sample period runs from 10/11/99, the date when EarthSat began to archive their forecasts electronically and make them publically available, through 10/22/01. Each weekday, EarthSat makes a set of h-day ahead daily average temperature forecasts, for $h = 1, 2, ..., 11$. EarthSat does not make forecasts on weekends.

We measure accuracy of all point forecasts as h-day-ahead root mean squared prediction error (RMSPE). We assess point forecasting accuracy at horizons of $h = 1, 2, ..., 11$ days, because those are the horizons at which EarthSat's forecasts are available, and we want to compare the performance of our

model to EarthSat's, among others

We also compute measures of the accuracy of our model and the EarthSat model relative to that of the persistence and climatological benchmarks. RMPSE ratios relative to benchmarks are known in the meteorological literature as skill scores (Brier and Allen, 1951) , and in the econometrics literature as U-statistics (Theil, 1966). Specifically, in an obvious notation, the skill score relative to the persistence forecast is

$$Skill_h^{pers} = \sqrt{\frac{\sum (\hat{y}_{t+h,t} - Y_{t+h})^2}{\sum (y_{t+h,t}^{pers} - Y_{t+h})^2}}, \tag{3a}$$

where $y_{t+h,t}^{pers} = y_t$ denotes the persistence forecast and $\hat{y}_{t+h,t}$ denotes either the autoregressive forecast or the EarthSat forecast, as desired. Similarly, the skill score relative to the climatological forecast is

$$Skill_h^{clim} = \sqrt{\frac{\sum (\hat{y}_{t+h,t} - Y_{t+h})^2}{\sum (\hat{y}_{t+h,t}^{clim} - Y_{t+h})^2}}, \tag{3b}$$

where $\hat{y}_{t+h,t}^{clim}$ denotes the climatological forecast, given by

$$\hat{y}_{t+h,t}^{clim} = \hat{\beta}_0 + \hat{\beta}_1(t+h) + \sum_{p=1}^{365} \hat{\delta}_i D_{i,t+h}. \tag{4}$$

A number of nuances merit discussion. First, for each of our time-series models, we estimate and forecast recursively, using only the data that were available in real time. Thus our forecasts at any time utilize no more average temperature information than do EarthSat's. In fact – and this is potentially important – our forecasts are based on *less* average temperature information: our forecast for day *t+1* made on day *t* is based on daily average temperature through 11:59 PM of day *t*, whereas the EarthSat forecast for day *t+1*, which is not released until 6:45 AM on day *t+1*, potentially makes use of the history of temperature through 6:45 AM of day *t+1*.

Second, we make forecasts using our models only on the dates that EarthSat made forecasts. In particular, we make no forecasts on weekends. Hence, our accuracy comparisons proceed by averaging squared errors over precisely the same days as those corresponding to the EarthSat errors. This ensures a fair apples-to-apples comparison.

We report RMSPEs in Table 4 at horizons of *h* = 1, 3, 5, 7, 9, and 11 days, for all cities and

forecasting models. In addition, we graph skill scores as a function of horizon, against the persistence forecast in Figure 13 and against the climatological forecast in Figure 14, for all cities and horizons. The results are the same for all cities, so it is not necessary to discuss them individually by city. The results most definitely do differ, however, across models and horizons, as we now discuss. We first discuss the performance of the time-series forecasts, and then we discuss the EarthSat forecasts.

Let us consider first the forecasting performance of the persistence, climatological, and autoregressive models across the various horizons. First consider the comparative performance of the persistence and climatological forecasts. When $h=1$, the climatological forecasts are much worse than the persistence forecasts, reflecting the fact that persistence in daily average temperature renders the persistence forecast quite accurate at very short horizons. As the horizon lengthens, however, this result is reversed: the persistence forecast becomes comparatively poor, as the temperature today has rather little to with the temperature, for example, nine days from now.

Second, consider the performance of the autoregressive forecasts relative to the persistence and climatological forecasts. Even when $h=1$, the autoregressive forecasts consistently outperform the persistence forecast, and their relative superiority *increases* with horizon. The autoregressive forecasts also outperform the climatological forecasts at short horizons, but their comparative superiority *decreases* with horizon. The performance of the autoregressive forecast is commensurate with that of the climatological forecast roughly by the time h=4, indicating that the cyclical dynamics captured by the autoregressive model, which are responsible for its superior performance at shorter horizons, are not very persistent and therefore not readily exploited for superior forecast performance at longer horizons.

Now let us compare the forecasting performance of the autoregressive model and the EarthSat model. When $h=1$, the EarthSat forecasts are much better than the autoregressive forecasts (which in turn are better then either the persistence or climatological forecasts, as discussed above). Figures 13 and 14 make clear, however, that the EarthSat forecasts outperform the autoregressive forecasts by progressively less as the horizon lengthens, with nearly identical performance obtaining by the time h=8.[22]

All told, we view the point forecasting results as encouraging. At short horizons, nonstructural time-series models produce forecasts at least as accurate, and often much more accurate, than those of standard benchmark persistence and climatological competitors, if nevertheless less accurate than state-of-the-art meteorological forecasts. At the longer horizons of relevance for weather derivatives,

[22] One could even make a case that the point forecasting performances of EarthSat and our three-component model become indistinguishable before h=8 (say, by h=5) if one were to account for the sampling error in the estimated RMSPEs and for the fact that the EarthSat information set for any day t actually contains a few hours of the next day.

nonstructural time-series models produce forecasts at least as accurate as those of its competitors. One might assert, of course, that doing no worse than competitors at long horizons is faint praise, as all point forecasts revert fairly quickly to the climatological forecast, and hence all long-horizon forecasts are arguably quite poor. But such an assertion misses a crucial point: modern time series methods can capture movements in conditional moments beyond the conditional mean, such as the conditional variance dynamics emphasized earlier, which feed into the distributional forecasts of ultimate relevance for weather derivatives, whose performance we now assess.

Distributional Forecasting

A important issue in weather forecasting for weather derivatives is how to use models that have been constructed for one horizon at other, longer, horizons, and how to produce not simply point forecasts, but distributional forecasts. For example, all of the models presented in this paper have been formulated, estimated and tested on daily data. In practice, however, weather derivatives are often used to hedge weather related risks over much longer horizons, such as six months, and what is required is not simply a point forecast over the long horizon, but rather an estimate of the entire conditional distribution of weather over that horizon, which is needed to price weather derivatives. In this section, we focus on such aggregation and distributional forecasting.

Before proceeding, we note that model aggregation may not always be an optimal strategy for modeling weather outcomes over varying horizons. In the presence of model misspecification, constructing a separate model for each horizon of interest may be a more suitable strategy. In the current setting, however, our goal is somewhat more general. Rather than providing the best model for each relevant horizon, here we are interested in understanding how well a daily model can approximate the distribution of weather outcomes at longer horizons. In this way, the analysis serves as a means of assessing the extent to which a single time-series model may be used in a variety of different forecasting contexts.

We use our model of daily average temperature to assess the distribution of cumulative *HDD*'s between November and March, for each city and for each year between 1960 and 2001:[23, 24]

[23] The summation runs from November 1 through March 31, for each of the 41 years in our sample. Because we removed February 29 from each leap year in our data set, each sum contains exactly 151 days.

[24] So as not to have an unnecessarily burdensome notation, we will often drop the y and I subscripts when the meaning is clear from context.

$$CumHDD_{y,i} = \sum_{t=1}^{151} HDD_{t,y,i} = \sum_{t=1}^{151} \max(65 - T_{t,y,i}, 0), \quad (5)$$

for y~=~1960,~ ..., ~2000, ~i~=~1, ~..., ~10. Note that the creation of *CumHDD* involves both nonlinear transformation of daily average temperature *T* into *HDD*, and then temporal aggregation of *HDD* into *CumHDD*.

We focus on *CumHDD* for two important reasons. First, weather derivative contracts are typically written on the cumulative sum of a weather related outcome, such as *HDD*s, *CDD*s, or rainfall, over a fixed horizon. Second, the November-March *HDD* contract is one of the most actively traded weather-related contracts and is of substantial interest to end users of weather models.

On October 31 of each year, and for each city, we use the estimated daily model to estimate the distribution of future *CumHDD* . We simulate 250 151-day realizations of *CumHDD*, which we then use to estimate the distribution, as follows. First, we simulate 250 151-day realizations of the temperature shock ε_t by drawing with replacement from the empirical distribution of estimated temperature shocks ($\hat{\varepsilon}_t$).[25, 26] Second, we run the 250 151-day realizations of temperature shocks through the estimated model (2) to obtain 250 simulated 151-day realizations of daily average temperature. Third, we convert the 250 simulated 151-day realizations of daily average temperature into 250 simulated 151-day realizations of *HDD*, which we cumulate over the November-March heating season,

$$CumHDD_s = \sum_{t=1}^{151} HDD_{t,s} = \sum_{t=1}^{151} \max(65 - T_{t,s}, 0), \; s = 1, 2, ..., 250. \quad (6)$$

Finally, we form the empirical distribution function of $CumHDD$, based upon $CumHDD_s$, s = 1, ..., 250.

After passing through the entire sample, we have 41 assessed distribution functions, $\hat{F}_y(\bullet), \; y = 1960, ..., 2000,$ one governing each of $CumHDD_y, \; y = 1960, ..., 2000.$ We investigate the conditional calibration of those distributional forecasts by examining the probability integral transform, as suggested and extended by Rosenblatt (1952), Dawid (1984), Diebold, Gunther and Tay (1998), and Diebold, Hahn and Tay (1999). In particular, if the estimated distribution and true distribution coincide year-by-year, then the series of assessed distribution functions $\hat{F}_y(\bullet)$ evaluated at the corresponding series of realized values of $CumHDD_y$, (denoted by $CumHDD_y^0$), should be distributed iid and uniformly on the

---

[25] It is important to use the empirical distribution of estimated shocks, because, as indicated in Figure 10, the distributions of standardized residuals typically display excess skewness and kurtosis.

[26] In an earlier analysis, we used a simulation sample of 1,000, and the differences in results were minimal.

unit interval. Formally, abstracting from parameter estimation uncertainty,

$$z_y \equiv \hat{F}_y(CumHDD_y^0) \overset{iid}{\sim} U(0,\ 1). \tag{7}$$

For each city, we check uniformity of z by examining histograms, and we check independence of z by examining correlograms of its first four powers. The sample of size 41 is of course small, but the framework has previously been applied successfully in small samples, as for example by Diebold, Tay and Wallis (1999).

First consider assessing uniformity. We estimate the density of z using simple four-bin histograms, which we present in the leftmost column of Figure 15, accompanied by 95% pointwise error bands under the *iid* $U(0, 1)$ null hypothesis.[27] Interestingly, the z series for the Midwestern cities in our sample, such as Chicago, Cincinnati, Dallas and Des Moines, appear approximately uniform, while those of eastern and western cities such as Philadelphia, Portland and Tucson, display greater deviations from uniformity.

The non-Midwestern deviations from uniformity, moreover, display a common pattern: too many large *CumHDD* realizations occur relative to the assessed distributions, and too few small realizations occur, as evidenced by the non-Midwestern cities' histograms increasing from left to right. The common nature of uniformity violations for non-Midwestern cities may indicate a common temperature component, due for example to El Niño, La Niña, changes in the jet stream, or various other global factors.

Now consider assessing the independence of the z series. We assess independence of the z series for each city via the correlograms of its first four powers, shown in the last four columns of Figure 15.[28] We display each correlogram with asymptotic 95% Bartlett bands under the *iid* null hypothesis. Interestingly, just as the Midwestern cities appeared to meet the uniformity criterion, so too do they appear to meet the *iid* criterion, with correlograms of all powers of z showing no significant deviation from white noise. In contrast, the non-Midwestern cities, taken as a whole, display greater divergence from *iid*, often corresponding to a common pattern of positive serial correlation at lags from one to three

[27] We note that the error bands provide exact coverage only if we abstract from parameter uncertainty associated with model estimation. Given the very large sample size considered here (more than 10,000), parameter uncertainty likely has little effect on the resulting inference.

[28] Given the sample size of 41, we computed correlograms to a maximum displacement of ten years.

years. Again, the common patterns in non-Midwestern deviations from *iid* may be due to dependence on one or more common factors.

In closing this section, let us elaborate on an important result: although our *CumHDD* distributional forecasting performance is encouraging, there is nevertheless clear room for improvement. Evidently the effects of small specification errors in the daily model, which have negligible consequences for near-term forecasting, cumulate as the horizon lengthens, producing large consequences for longer-term forecasting. This makes perfect sense when one considers that the error in forecasting *CumHDD* is of course the sum of the many component errors, and that the variance of a sum is the sum of the variances *plus* the sum of all possible pairwise covariances. Hence tiny and hard-to-detect but slowly-decaying serial correlation in 1-day-ahead daily average temperature forecasting errors may significantly inflate the variance of *CumHDD* over long horizons.

5. Summary, Concluding Remarks and Directions for Future Research

Weather modeling and forecasting are crucial to both the demand and supply sides of the weather derivatives market. On the demand side, to assess the potential for hedging against weather surprises and to formulate the appropriate hedging strategies, one needs to determine how much "weather noise" exists for weather derivatives to eliminate, and that requires weather modeling and forecasting. On the supply side, standard approaches to arbitrage-free pricing are irrelevant in weather derivative contexts, and so the only way to price options reliably is again by modeling and forecasting the underlying weather variable. Rather curiously, it seems that little thought has been given to the crucial question of how best to approach weather modeling and forecasting in the context of weather derivative demand and supply. The vast majority of extant weather forecasting literature has a structural "atmospheric science" feel, and although such an approach may well be best for forecasting twelve hours ahead, as verified both by our own results and those of others, it is not obvious that it is best for the longer horizons relevant for weather derivatives, such as twelve weeks or six months. Moreover, it is distributional forecasts, not point forecasts, that are of maximal relevance in the derivatives context. Good distributional forecasting does not necessarily require a structural model, but it does require accurate approximations to stochastic dynamics.

In this paper we took an arguably-naive nonstructural time-series approach to modeling and forecasting daily average temperature in ten U.S. cities, and we inquired systematically as to whether it proves useful. The answer, perhaps surprisingly, is yes. Time series modeling reveals a wealth of information about both conditional mean dynamics and conditional variance dynamics of daily average temperature, and it provides insights into both the distributions of temperature and temperature surprises,

and the differences between them. The success of time-series modeling in capturing conditional mean dynamics translates into successful point forecasting. Our point forecasts dominated the persistence and climatological forecasts, but were still not as good as the judgementally-adjusted NWP forecast produced by EarthSat until a horizon of eight days, after which all point forecasts perform equally well.

We routinely found strong seasonality in weather surprise volatility, and we assessed the adequacy of long-horizon distributional forecasts that accounted for it, with mixed but encouraging results. Interestingly, we found that adequacy of conditional calibration varies regionally, indicating possible dependence on common latent components, perhaps due to El Niño or La Niña.

All told, we would assert that in the context of weather modeling as relevant for weather derivatives, it appears that simple yet sophisticated time series models and forecasts provide an easily-traversed route to the well-calibrated long-horizon distributional forecasts of maximal relevance to weather derivatives. We would also assert that our views are consistent with the mainstream consensus in atmospheric science. In his well-known text, for example, Wilks (1995, p. 159) notes that

> [Statistical weather forecasting] methods are still viable and useful at very short lead times (hours in advance) or very long lead times (weeks or more in advance) for which NWP information is either not available with sufficient promptness or accuracy, respectively.

Indeed, in many respects our results are simply an extensive confirmation of Wilks' assertion in the context of weather derivatives, which are of great current interest. When, in addition, one considers that time series models and methods are inexpensive, easily replicated, easily extended, beneficially intrinsically stochastic, and capable of producing both point and density forecasts at a variety of horizons, we feel that a strong case exists for their use in the context of modeling and forecasting as relevant for weather derivatives.

Ultimately, our view on weather forecasting for weather derivatives has evolved toward recognition that climatological forecasts are what we need, but that standard point climatological forecasts – effectively little more than daily averages – are much too restrictive. Instead, we seek "generalized climatological forecasts" from richer models tracking entire conditional distributions, and time series methods may have much to contribute. Our model quantifying the time-varying conditional variance of daily average temperature is one step toward a fully generalized climatological forecast, but many issues remain unexplored. Here we list a few of those that we find most intriguing.

Richer Climatological Dynamics

Arguably our central modeling innovation centers on the dynamics of the conditional variance.

But richer dynamics could also be beneficially permitted in both lower-ordered conditional moments (the conditional mean) and higher-ordered conditional moments (such as the conditional skewness and kurtosis).

As regards the conditional mean, one could introduce explanatory variables, as in Visser and Molenaar (1995), who condition on a volcanic activity index, sunspot numbers, and a southern oscillation index. Relevant work also includes Jones (1996) and Pozo et al. (1998), but all of those papers use annual data and therefore miss the seasonal patterns in both conditional mean and conditional variance dynamics so crucial for weather derivatives demand and supply. One could also allow for stochastic regime switching, which might aid, for example, in the detection of El Niño and La Niña events, as in Richman and Montroy (1996).[29]

As regards the conditional skewness and kurtosis, one could model them directly, as for example with the autoregressive conditional skewness model of Harvey and Siddique (1999). Alternatively, one could directly model the evolution of the entire conditional density, as in Hansen (1994).

Forecasting Under the Relevant Loss Function

We have already imbedded the relevant loss function in our analysis, at least in part, by modeling directly the object of interest (daily average temperature) as opposed to its underlying components (daily maximum and minimum temperature). One could go farther, however, by fitting different forecasting models to average temperature for use at different horizons. Presently we fit only a single (daily) average temperature model, which we estimate by minimizing a loss function corresponding to 1-day-ahead mean squared prediction error, and we then use the model for forecasting at all horizons. Alternatively, we could estimate by minimizing h-day-ahead mean squared prediction error when forecasting h days ahead, effectively fitting customized models at each horizon.

Direct Modeling of Daily Maximum and Minimum Temperatures

Thus far we have adopted an aggregated, or "top-down," approach, simply modeling and forecasting daily average temperature directly. We believe that there are good arguments for doing so, linked to the possibility of misspecification, as discussed earlier. It may nevertheless be of some interest also to explore a disaggregated, or "bottom- up," approach, modeling and forecasting separately the daily minimum and maximum, and then averaging their separate forecasts. Even under correct specification, perhaps surprisingly, it can not be ascertained a priori whether the aggregated or disaggregated approach

[29] See also Zwiers and von Storch (1990). Note that allowance for stochastic regime switching is very different from structural break detection, caused, for example, by a change in measurement location, measurement equipment, observation time.

is better for forecasting. The answer depends on the specifics of the situation; the only way to tell is to try both approaches and compare the forecasting results, as shown in Pesaran, Pierse and Kumar (1989).

Formal Unobserved Components Modeling

It would be interesting to contrast the forecasting performance of the simple components models used here with that of formal unobserved-components models a la Harvey (1989) potentially involving stochastic trend and/or seasonality, in contrast to the deterministic trends and seasonals used in the present paper.

Multivariate Analysis and Cross-Hedging

Cross-city correlations may be crucially important, because they govern the potential for cross-hedging. Hedging weather risk in a remote Midwestern location might, for example, be prohibitively expensive or even impossible due to illiquid or nonexistent markets, but if that risk is highly correlated with Chicago's weather risk, for which a liquid market exists, effective hedging may still be possible. An obvious and important extension of the univariate temperature analysis reported in the present paper is vector autoregressive modeling of daily average temperature in a set of cities, with multivariate GARCH disturbances (note that conditional covariance may be time-varying just as are the conditional variances). Of particular interest will be the fitted and forecasted conditional mean, variance and covariance dynamics, the covariance matrices of standardized innovations, and the impulse response functions (which chart the speed and pattern with which weather surprises in one location are transmitted to other locations).

Weather, Earnings and Share Prices

An interesting multivariate issue involves weather-related swings in earnings. One might conjecture that once Wall Street recognizes that there are effective ways to manage weather risk, weather-related swings in earnings will no longer be tolerated. It will be of interest to use the size of weather-related swings in earnings as way to assess the potential for weather derivatives use. In particular, we need to understand how weather surprises translate into earnings surprises, which then translate into stock return movements. Some interesting subtleties may arise. As one example, note that only systematic weather risk should be priced, which raises the issue of how to disentangle systematic and non-systematic weather risks. As a second example, note that there may be nonlinearities in the relationship between prices and the weather induced via path dependence; if there is a freeze early-on, it doesn't matter how good the weather is subsequently: the crop will be ruined and prices will be high (Richardson, Bodoukh, Sjen, and Whitelaw, 2001).

References

Andersen, T.G., Bollerslev, T. and Diebold, F.X. (2003), "Parametric and Nonparametric Volatility Measurement," forthcoming in L.P. Hansen and Y. Ait-Sahalia (eds.), *Handbook of Financial Econometrics*. Amsterdam: North-Holland. Available at www.ssc.upenn.edu/~fdiebold.

Bloomfield, P. (1992), "Trends in Global Temperature," *Climate Change*, 21, 1-16.

Brier, G.W. and Allen, R.A. (1951), "Verification of Weather Forecasts," in T.F. Malone (ed.), *Compendium of Meteorology*, American Meteorological Society.

Cao, M. and Wei, J. (2001), "Pricing Weather Derivatives: An Equilibrium Approach," Manuscript, University of York and University of Toronto.

Christoffersen, P.F. and Diebold, F.X. (1996), "Further Results on Forecasting and Model Selection under Asymmetric Loss," *Journal of Applied Econometrics*, 11, 561-572.

Christoffersen, P.F., and Diebold, F.X. (1997), "Optimal Prediction Under Asymmetric Loss," *Econometric Theory*, 13, 808-817.

Davis, M. (2001), "Pricing Weather Derivatives by Marginal Value," *Quantitative Finance*, 1, 305-308.

Dawid, A.P. (1984), "Statistical Theory: The Prequential Approach," *Journal of the Royal Statistical Society*, Series A, 147, 278-292.

Diebold, F.X. (2001), *Elements of Forecasting*, Second Edition. Cincinnati: South-Western College Publishing.

Diebold, F.X., Gunther, T. and Tay, A.S. (1998), "Evaluating Density Forecasts, with Applications to Financial Risk Management," *International Economic Review*, 39, 863-883.

Diebold, F.X., Hahn, J. and Tay, A.S. (1999), "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-Frequency Returns on Foreign Exchange," *Review of Economics and Statistics*, 81, 661-673.

Diebold, F.X., Tay, A.S. and Wallis, K. (1999), "Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters," in R. Engle and H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, 76-90. Oxford: Oxford University Press.

Engle, R.F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987-1008.

Geman, H., Ed. (1998), *Insurance and Weather Derivatives: From Exotic Options to Exotic Underlyings*. London: Risk Publications.

Hansen, B.E. (1994), "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705-730.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, C.R. and Siddique, A. (1999), "Autoregressive Conditional Skewness," *Journal of Financial and Quantitative Analysis*, 34, 465-488.

Hyndman, R.J. and Grunwald, G.K. (2000), "Generalized Additive Modeling of Mixed Distribution Markov Models with Application to Melbourne's Rainfall," *Australian and New Zealand Journal of Statistics*, 42, 145-158.

Jones, R.H. (1996), "The Potential of State Space Models in Statistical Meteorology," *Proceedings of the 13th Conference on Probability and Statistics in the Atmospheric Sciences*, American Meteorological Society.

Milionis, A.E. and Davies, T.D. (1994), "Box-Jenkins univariate Modeling for Climatological Time Series Analysis: An Application to the Monthly Activity of Temperature Inversions," *International Journal of Climatology*, 14 569-579.

Neese, J.M. (1994), "Systematic Biases in Manual Observations of Maximum and Minimum Temperature," *Journal of Climate*, 7, 834-842.

Pesaran, M.H., Pierse, R.G. , Kumar, M.S. (1989), "Econometric Analysis of Aggregation in the Context of Linear Prediction Models," *Econometrica*, 57, 861-888.

Pozo, D., Esteban-Parra, M.J., Rodrigo, F.S. and Castro-Diez, Y. (1998), "ARIMA Structural Models for Temperature Time Series of Spain," *Proceedings of the 14th Conference on Probability and Statistics in the Atmospheric Sciences*, American Meteorological Society.

Richardson, M., Bodoukh, J., Sjen, Y. and Whitelaw, R. (2001), "Freshly Squeezed: A Reinterpretation of Market Rationality in the Orange Juice Futures Market," Working Paper, New York University.

Richman, M.B. and Montroy, D.L. (1996), "Nonlinearities in the Signal Between El Niño / La Niña Events and North American Precipitation and Temperature," *Proceedings of the 13th Conference on Probability and Statistics in the Atmospheric Sciences*, American Meteorological Society.

Roll, R. (1984), "Orange Juice and Weather," *American Economic Review*, 74, 861-880.

Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, 23, 470-472.

Seater, J.J. (1993), "World Temperature-Trend Uncertainties and Their Implications for Economic Policy," *Journal of Business and Economic Statistics*, 11, 265-277.

Storch, H. von and Zwiers, F.W. (1999), *Statistical Analysis in Climate Research*. Cambridge: Cambridge University Press.

Theil, H. (1966), *Applied Economic Forecasting*. Amsterdam: North-Holland.

Torro, H., Meneu, V. and Valor E. (2001), "Single Factor Stochastic Models with Seasonality Applied to Underlying Weather Derivatives Variables," Manuscript, University of Valencia.

Tribia, J.J. (1997), "Weather Prediction," in R. Katz and A. Murphy (eds.), *Economic Value of Weather and Climate Forecasts*. Cambridge: Cambridge University Press.

Visser, H. and Molenaar, J. (1995), "Trend Estimation and Regression Analysis in Climatological Time Series: An Application of Structural Time Series Models and the Kalman Filter," *Journal of Climate*, 8, 969-979.

Weather Risk (1998), Supplement to *Risk Magazine*, October.

Weather Risk (2000), Supplement to *Risk Magazine*, August.

Wilks, D.S. (1995), *Statistical Methods in the Atmospheric Sciences*. New York: Academic Press.

Zwiers, F.W. and von Storch, H. (1990), "Regime Dependent Autoregressive Time Series Modeling of the Southern Oscillation," *Journal of Climate*, 3, 1347-1363.

Table 1
Temperature Measuring Stations

City	State	Measuring Station	Measuring Station Symbol
Atlanta	GA	Hartsfield Airport	ATL
Chicago	IL	O'Hare Airport	ORD
Cincinnati	OH	Covington, KY	CVG
Dallas	TX	Dallas - Fort Worth	DFW
Des Moines	IA	Des Moines Int'l Airport	DSM
Las Vegas	NV	McCarran Int'l Airport	LAS
New York	NY	La Guardia	LGA
Philadelphia	PA	Philadelphia Int'l Airport	PHL
Portland	OR	Portland Int'l Airport	PDX
Tucson	AZ	Tucson Airport	TUS

Table 2
Temperature-Related Variables

Variable	Definition
TMAX	Daily Maximum Temperature - rounded to the nearest integer
TMIN	Daily Minimum Temperature - rounded to the nearest integer
T	Daily Average Temperature, (*TMAX*+*TMIN*)/2
HDD	Daily Heating Degree Days, max(0, 65-*T*)
CDD	Daily Cooling Degree Days, max(0, *T*-65)

Table 3
Missing Data

Date	Measuring Station	Variable	Reference Station	30 Year Average Difference	σ(*Difference*)
9/15/96	DSM	*TMAX*	Omaha, NE	-0.33	4.40
9/15/96	DSM	*TMIN*	Omaha, NE	0.83	3.93
5/1/00	PHL	*TMAX*	Allentown, PA	1.70	2.33
5/1/00	PHL	*TMIN*	Allentown, PA	4.27	3.60
5/20/00	LGA	*TMIN*	New York CP, NY	0.40	1.87
6/23/00	PDX	*TMIN*	Salem, OR	4.83	2.83
7/12/00	PHL	*TMAX*	Allentown, PA	2.60	4.22
7/19/00	LGA	*TMAX*	New York CP, NY	-1.03	1.87
7/20/00	LGA	*TMAX*	New York CP, NY	-0.30	2.20
9/15/00	CVG	*TMAX*	Columbus, OH	1.30	2.89
10/3/00	PHL	*TMIN*	Allentown, PA	4.83	3.68
10/4/00	PHL	*TMIN*	Allentown, PA	4.43	3.08

Notes: The table contains information regarding the weather stations used when the primary weather station records a missing value. The table lists the date the missing value was recorded, the measuring station recording a missing value, the missing variable and the location of the reference station used in place of the ususal measuring station. The table also reports the thirty-year mean difference in the missing variable as well as the thirty-year standard deviation of the difference.

Table 4
Point Forecast Accuracy Comparisons
Daily Average Temperature

	Persistence	Climatological	Autoregressive	EarthSat
Atlanta				
1-Day-Ahead	4.50	6.93	4.12	2.74
3-Day-Ahead	8.00	6.88	6.45	3.84
5-Day-Ahead	8.72	6.84	6.69	5.10
7-Day-Ahead	9.07	7.04	7.03	6.04
9-Day-Ahead	8.99	6.93	6.89	6.65
11-Day-Ahead	9.28	6.59	6.59	7.00
Chicago				
1-Day-Ahead	6.73	8.74	6.06	3.22
3-Day-Ahead	10.50	8.72	8.38	4.70
5-Day-Ahead	11.06	8.72	8.57	6.31
7-Day-Ahead	11.54	8.50	8.45	7.46
9-Day-Ahead	11.74	8.88	8.84	8.48
11-Day-Ahead	11.99	8.55	8.53	8.92
Cincinnati				
1-Day-Ahead	6.25	8.52	5.61	2.96
3-Day-Ahead	10.30	8.58	8.21	4.49
5-Day-Ahead	11.29	8.67	8.56	6.28
7-Day-Ahead	11.54	8.50	8.45	7.46
9-Day-Ahead	11.46	8.55	8.52	8.34
11-Day-Ahead	11.82	8.51	8.51	9.09
Dallas				
1-Day-Ahead	5.87	7.67	5.35	3.33
3-Day-Ahead	9.07	7.69	7.26	4.79
5-Day-Ahead	9.18	7.81	7.52	6.00
7-Day-Ahead	9.90	8.00	7.88	6.83
9-Day-Ahead	10.13	7.57	7.53	7.55
11-Day-Ahead	10.11	7.65	7.63	8.27
Des Moines				
1-Day-Ahead	6.97	9.50	6.37	4.00
3-Day-Ahead	10.41	9.59	8.58	5.18
5-Day-Ahead	11.18	9.48	9.08	6.70
7-Day-Ahead	11.42	9.52	9.31	8.30
9-Day-Ahead	12.43	9.47	9.40	8.93
11-Day-Ahead	12.66	9.27	9.23	9.55

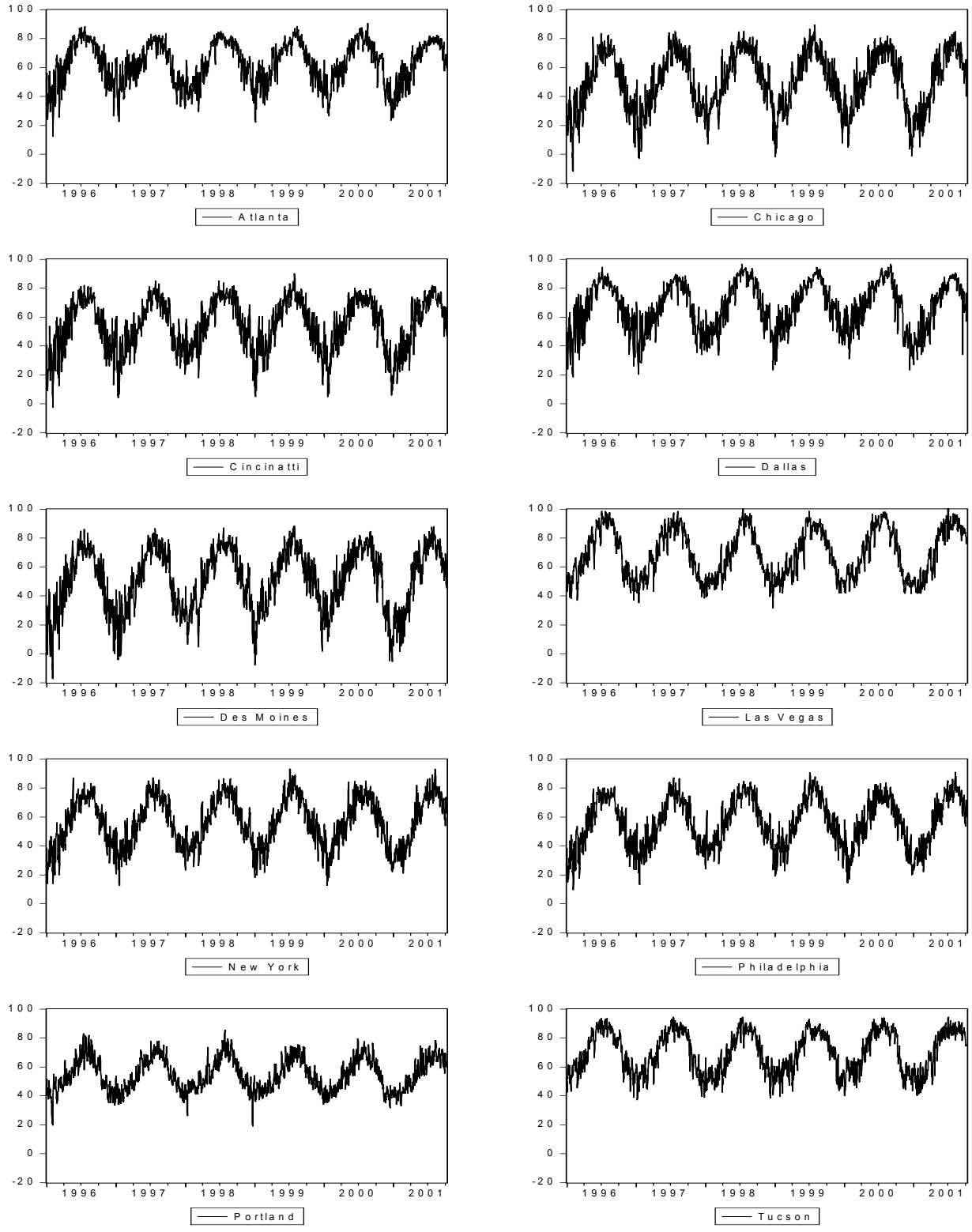
Notes: We show each forecast's root mean squared error, measured in degrees Fahrenheit.

Table 4 (Continued)
Point Forecast Accuracy Comparisons
Daily Average Temperature

	Persistence	Climatological	Autoregressive	EarthSat
Las Vegas				
1-Day-Ahead	3.78	5.99	3.57	2.54
3-Day-Ahead	6.15	5.85	5.20	3.28
5-Day-Ahead	7.08	5.80	5.58	4.19
7-Day-Ahead	7.71	6.02	5.92	5.32
9-Day-Ahead	7.96	5.97	5.97	5.81
11-Day-Ahead	7.93	5.80	5.78	6.04
New York				
1-Day-Ahead	5.84	7.22	5.23	2.70
3-Day-Ahead	9.17	6.93	6.84	4.11
5-Day-Ahead	10.16	7.18	7.19	5.63
7-Day-Ahead	9.85	7.19	7.16	6.50
9-Day-Ahead	9.58	7.18	7.15	7.53
11-Day-Ahead	10.27	6.98	6.98	8.39
Philadelphia				
1-Day-Ahead	5.53	7.12	4.95	2.61
3-Day-Ahead	8.83	6.95	6.74	3.91
5-Day-Ahead	9.83	7.27	7.23	5.35
7-Day-Ahead	9.87	7.19	7.15	6.26
9-Day-Ahead	9.55	6.98	6.95	7.24
11-Day-Ahead	10.18	7.19	7.08	8.37
Portland				
1-Day-Ahead	3.71	4.57	3.34	2.49
3-Day-Ahead	5.85	4.78	4.67	3.47
5-Day-Ahead	3.71	4.57	3.34	2.49
7-Day-Ahead	6.58	4.70	4.70	4.52
9-Day-Ahead	6.73	4.75	4.75	5.16
11-Day-Ahead	6.60	4.72	4.72	5.37
Tucson				
1-Day-Ahead	3.81	6.04	3.52	3.39
3-Day-Ahead	6.44	5.89	5.36	4.04
5-Day-Ahead	7.47	5.75	5.62	4.79
7-Day-Ahead	7.88	6.06	5.99	5.29
9-Day-Ahead	8.05	6.04	6.01	5.46
11-Day-Ahead	7.49	5.72	5.69	5.70

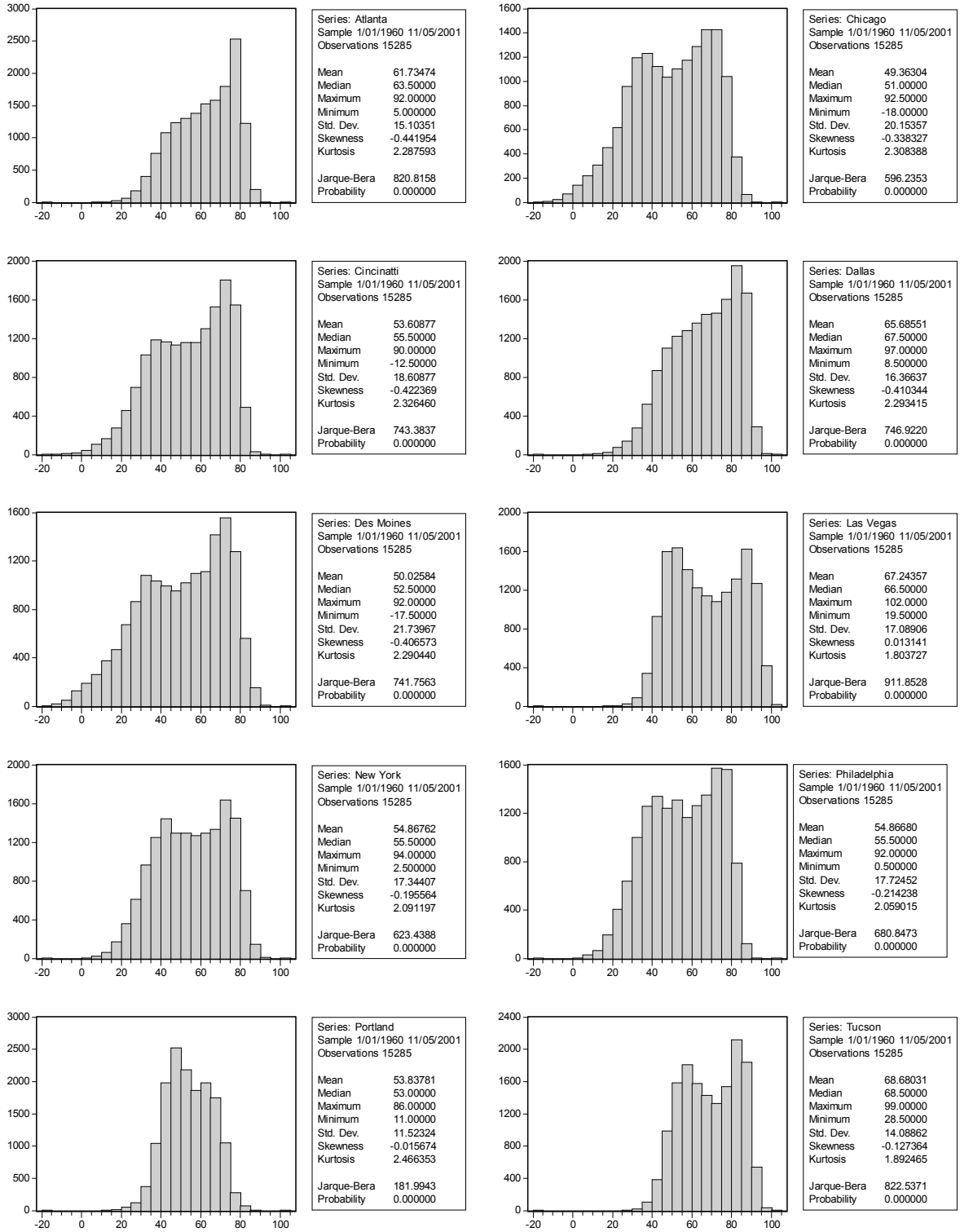
Notes: We show each forecast's root mean squared error, measured in degrees Fahrenheit.

Figure 1
Time Series Plots of Daily Average Temperature



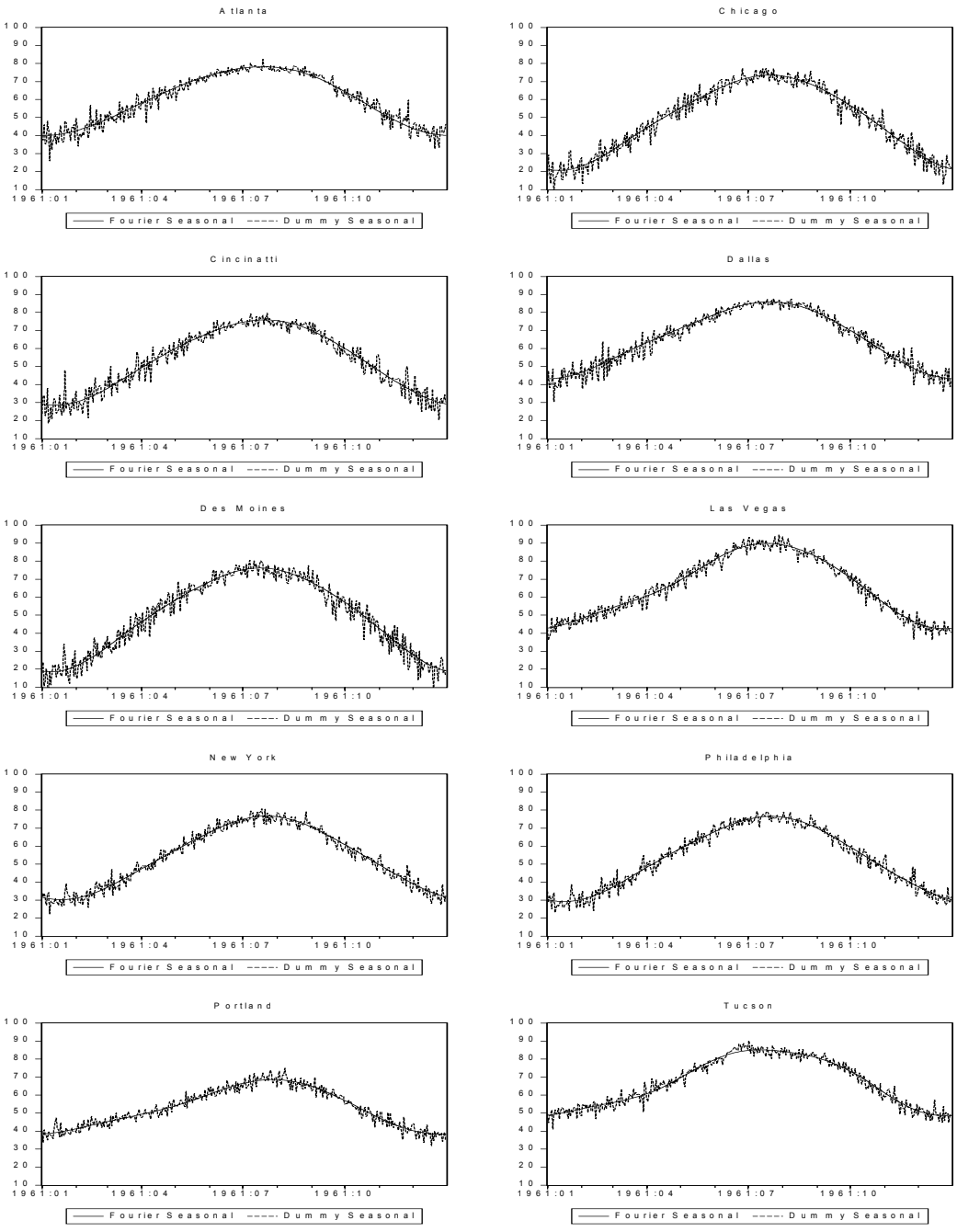
Notes: Each panel displays a time series plot of daily average temperature, 1996-2001.

Figure 2
Estimated Unconditional Distributions of Daily Average Temperature



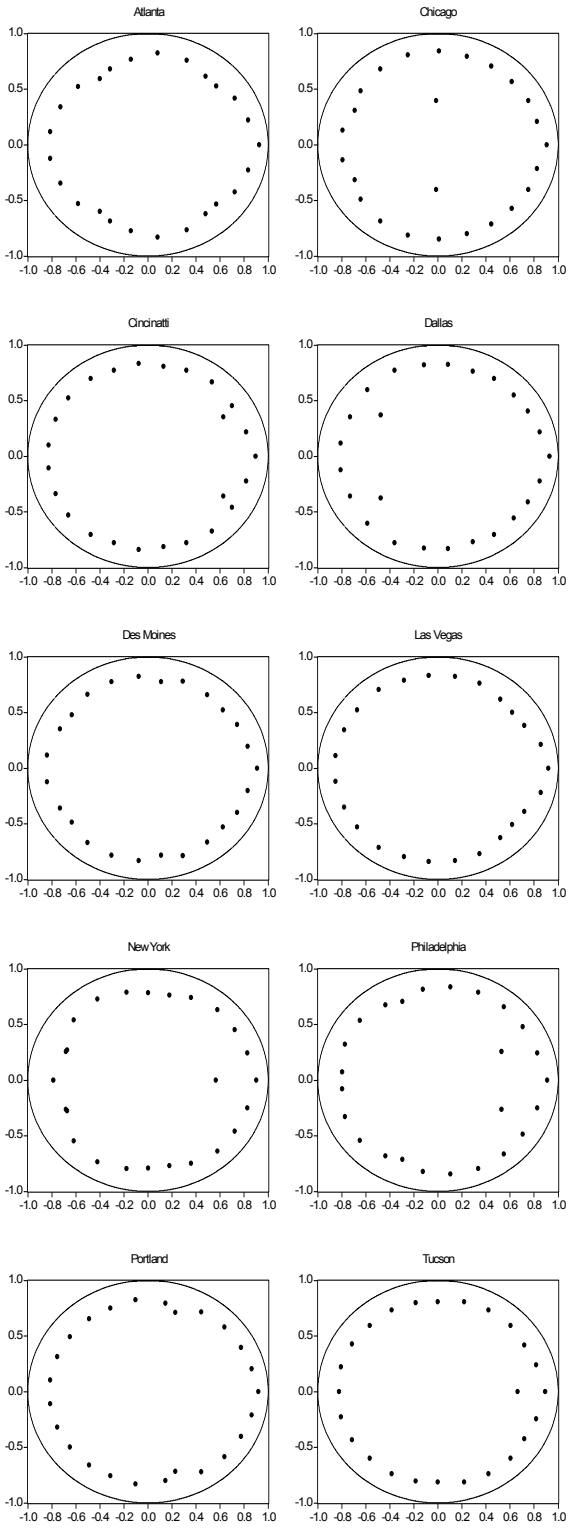
Notes: Each panel displays a histogram and associated descriptive statistics of daily average temperature, 1960-2001.

Figure 3
Estimated Seasonal Patterns of Daily Average Temperature
Fourier Series vs. Daily Dummies



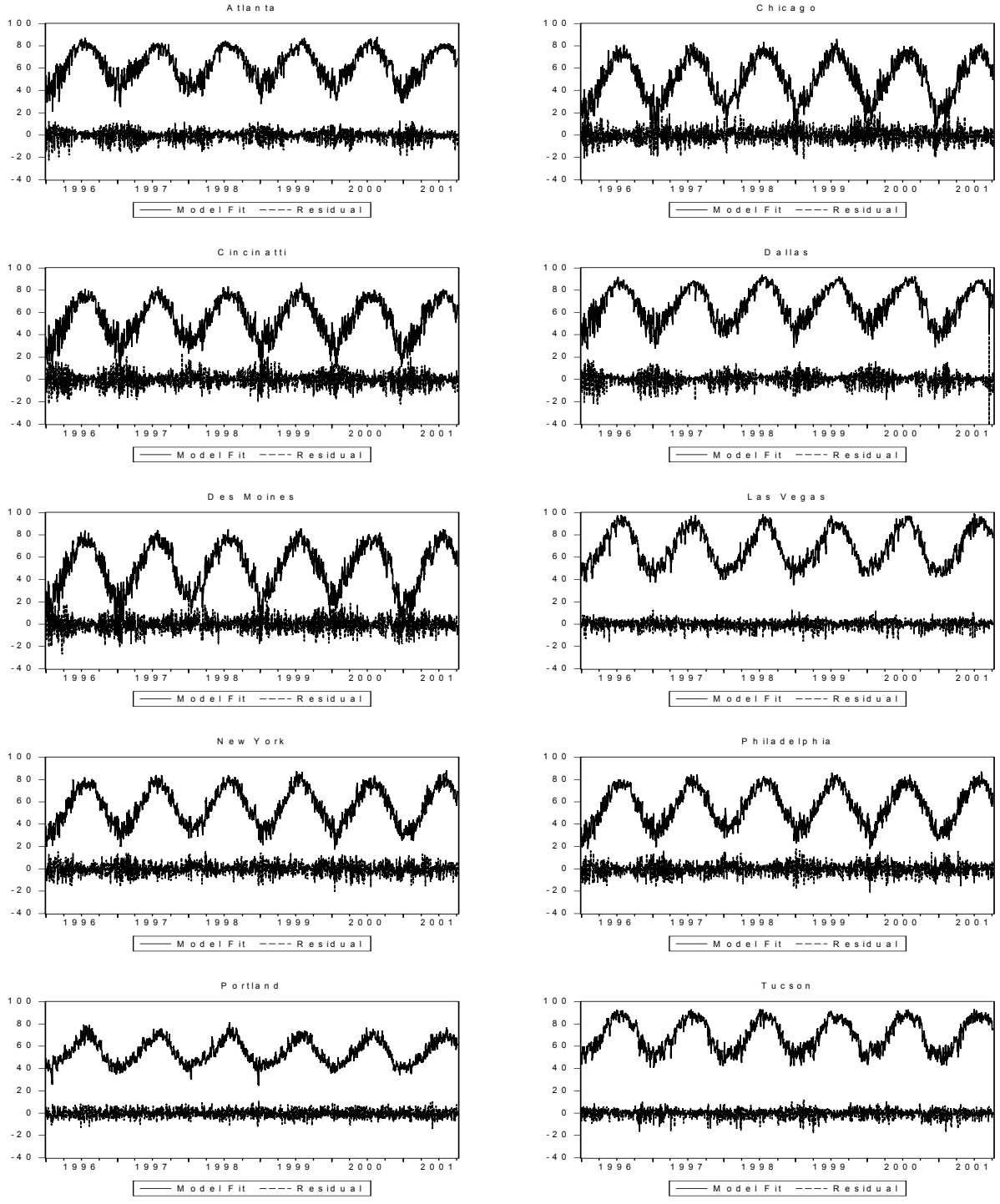
Notes: We show smooth seasonal patterns estimated from Fourier models, $Seasonal_t = \sum_{p=1}^{3} \left(\sin\left(2\pi p \frac{n(t)}{365}\right) + \beta_i \cos\left(2\pi p \frac{n(t)}{365}\right) \right)$, and rough seasonal patterns estimated from dummy variable models, $Seasonal_t = \sum_{i=1}^{365} \delta_i D_{it}$, as discussed in detail in the text.

Figure 4
Inverted Estimated Autoregressive Roots
of Daily Average Temperature



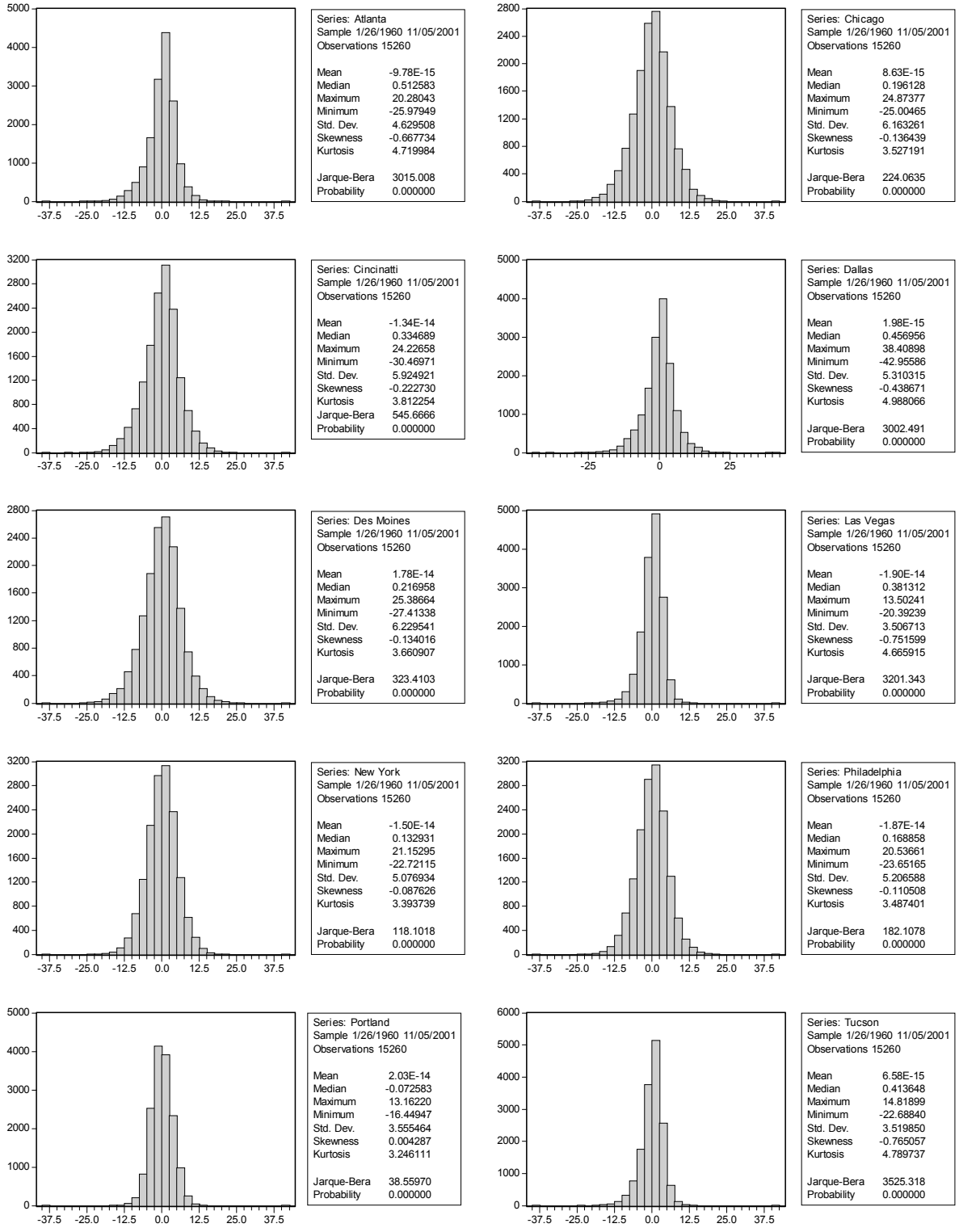
Notes: Each panel displays the reciprocal roots of the autoregressive lag operator polynomial, $1-\sum_{l=1}^{25}\hat{\rho}_{t-l}B^{l}$, associated with the estimated model $\hat{T}_t = \hat{Trend}_t + \hat{Seasonal}_t + \sum_{l=1}^{25}\hat{\rho}_{t-l}T_{t-l} + \hat{\sigma}_t\varepsilon_t$, as discussed in detail in the text. The real component of each inverse root appears on the horizontal axis, and the imaginary component appears on the vertical axis.

Figure 5
Actual Values, Fitted Values, and Residuals
Daily Average Temperature



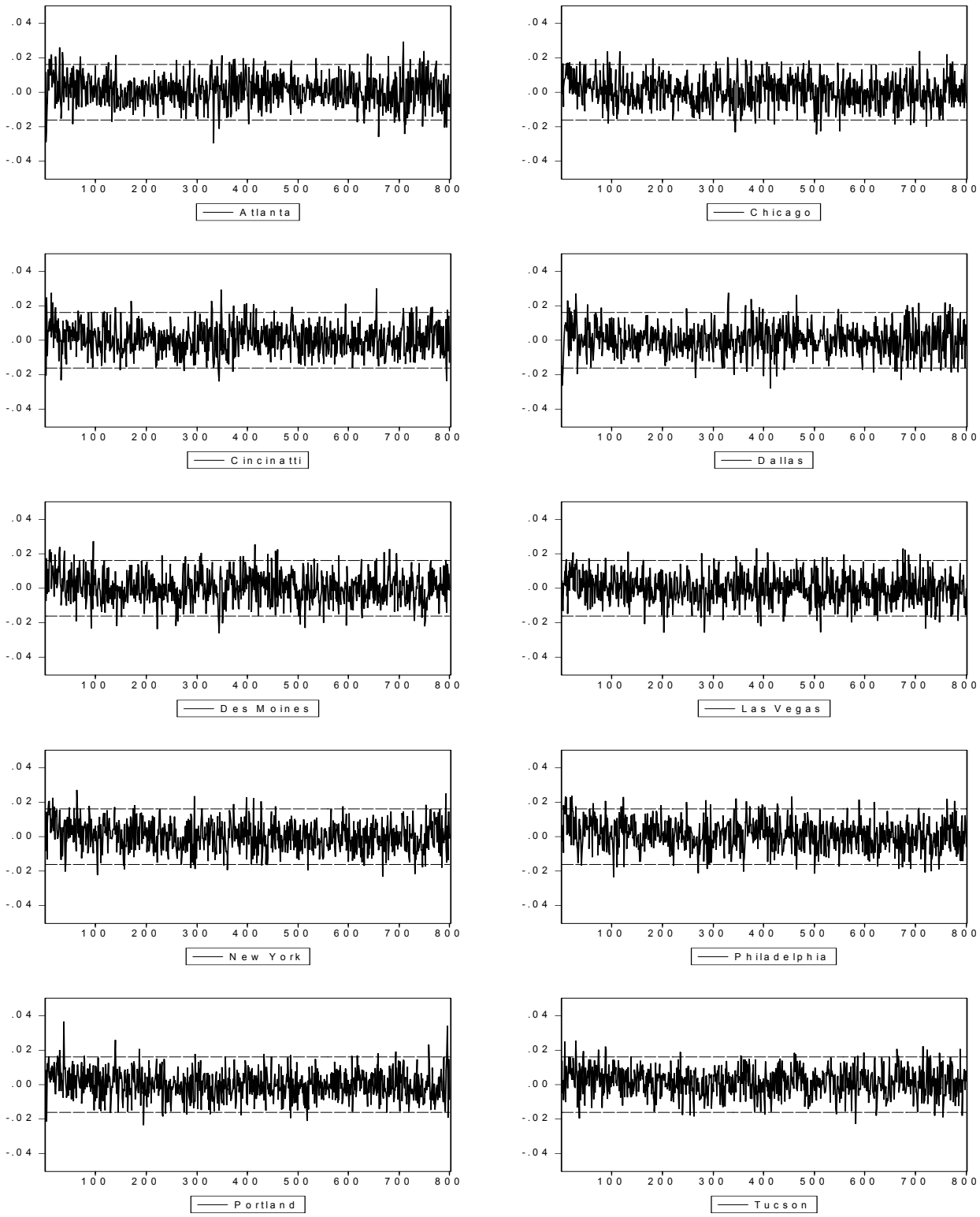
Notes: Each panel displays actual values, fitted values, and residuals from an unobserved-components model, $T_t = Trend_t + Seasonal_t + \sum_{i=1}^{25} \rho_i T_{t-i} + \sigma_t \varepsilon_t$, as discussed in detail in the text, 1996-2001.

Figure 6
Estimated Unconditional Distributions of Residuals
Daily Average Temperature



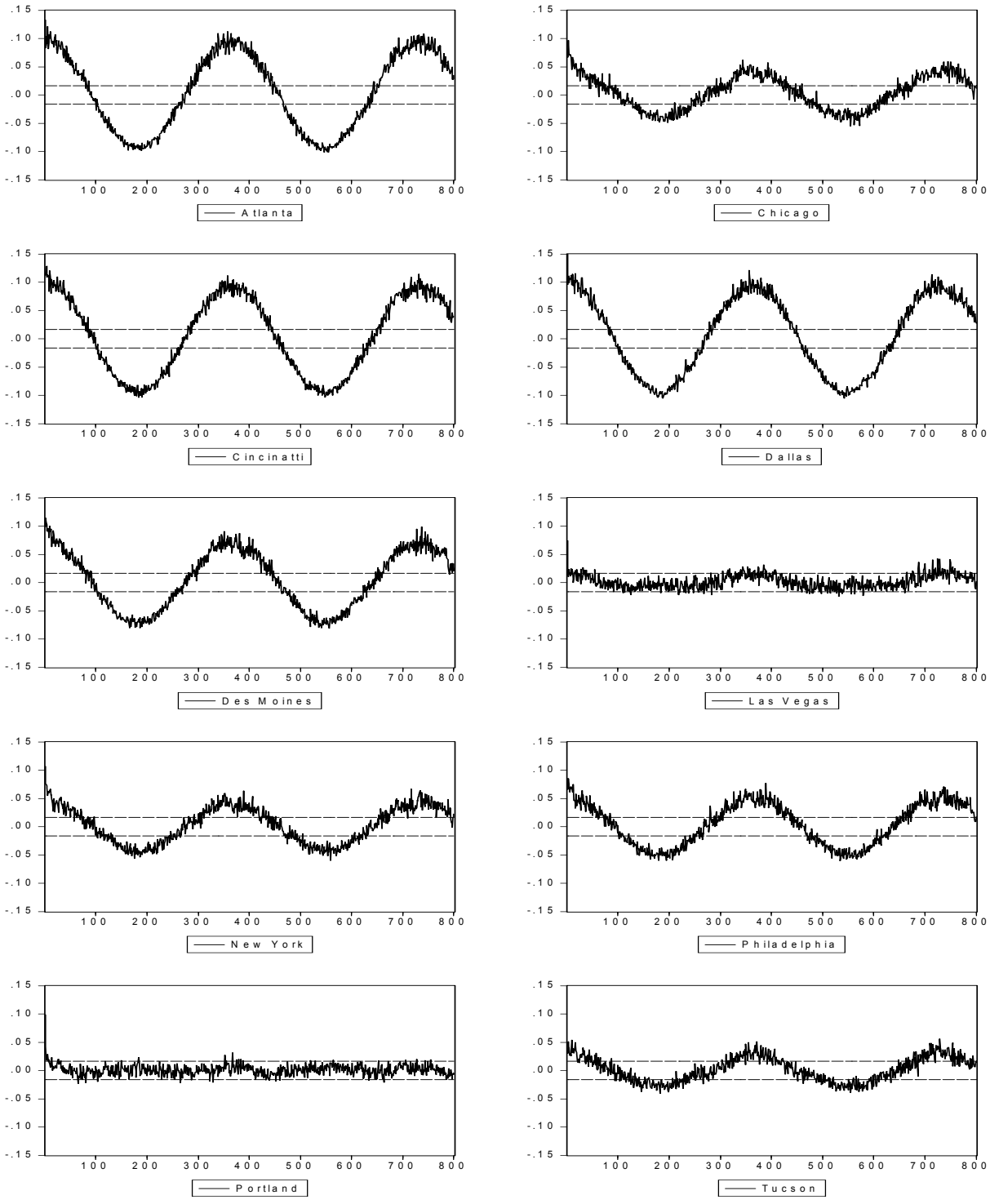
Notes: Each panel displays an empirical distribution and related statistics for the residuals from our daily average temperature model, $T_t - \hat{T}rend_t - \hat{S}easonal_t - \sum_{i=1}^{25} \hat{\rho}_i T_{t-i}$, as discussed in detail in the text.

Figure 7
Correlograms of Residuals
Daily Average Temperature



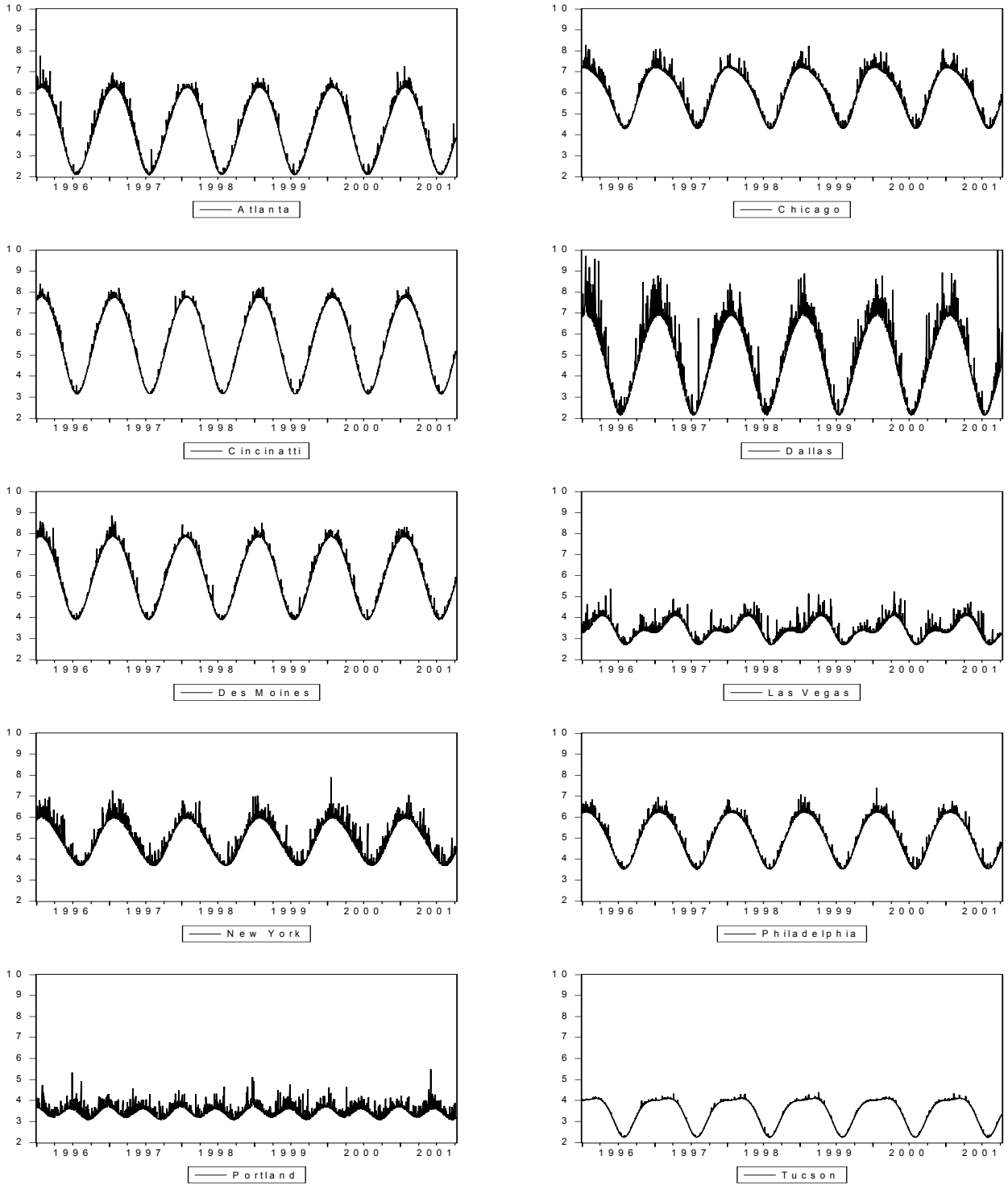
Notes: Each panel displays sample autocorrelations of the residuals from our daily average temperature model, $T_t - \hat{T}rend_t - \hat{S}easonal_t - \sum_{i=1}^{25} \hat{\rho}_i T_{t-i}$, together with Bartlett's approximate ninety-five percent confidence intervals under the null hypothesis of white noise, as discussed in detail in the text.

Figure 8
Correlograms of Squared Residuals
Daily Average Temperature



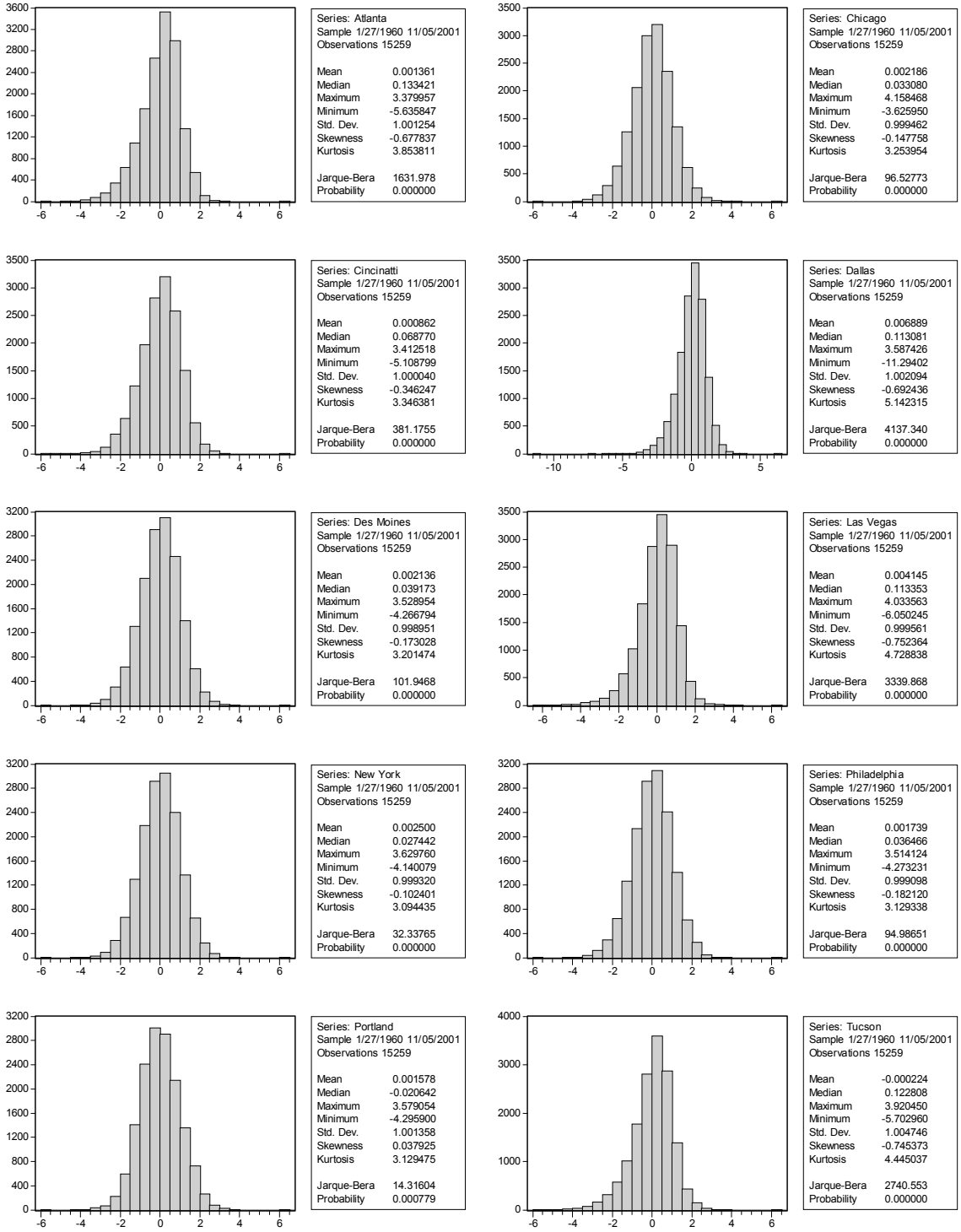
Notes: Each panel displays sample autocorrelations of the squared residuals from our daily average temperature model, $\left(T_t - \hat{T}rend_t - \hat{S}easonal_t - \sum_{i=1}^{25} \hat{\rho}_i T_{t-i}\right)^2$, together with Bartlett's approximate ninety-five percent confidence intervals under the null hypothesis of white noise, as discussed in detail in the text.

Figure 9
Estimated Conditional Standard Deviations
Daily Average Temperature



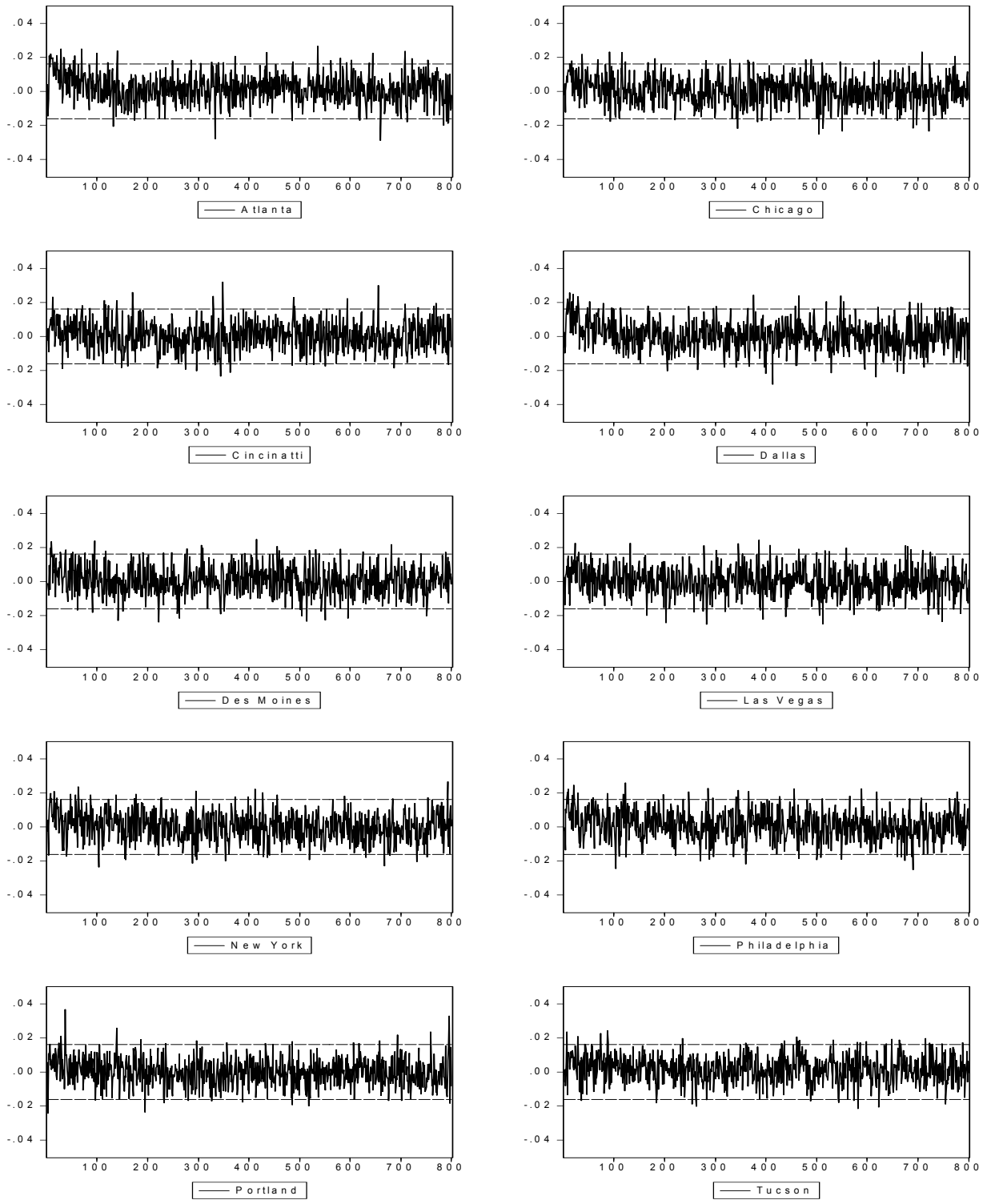
Notes: Each panel displays a time series of estimated conditional standard deviations of daily average temperature obtained from the model, $\hat{\sigma}_t = \sum_{q=1}^{2} \hat{\gamma}_{c,q}\cos\left(2\pi q\frac{n(t)}{365}\right) + \hat{\gamma}_{s,q}\sin\left(2\pi q\frac{n(t)}{365}\right) + \hat{\alpha}\varepsilon_{t-1}^2$.

Figure 10
Estimated Unconditional Distributions of Standardized Residuals
Daily Average Temperature



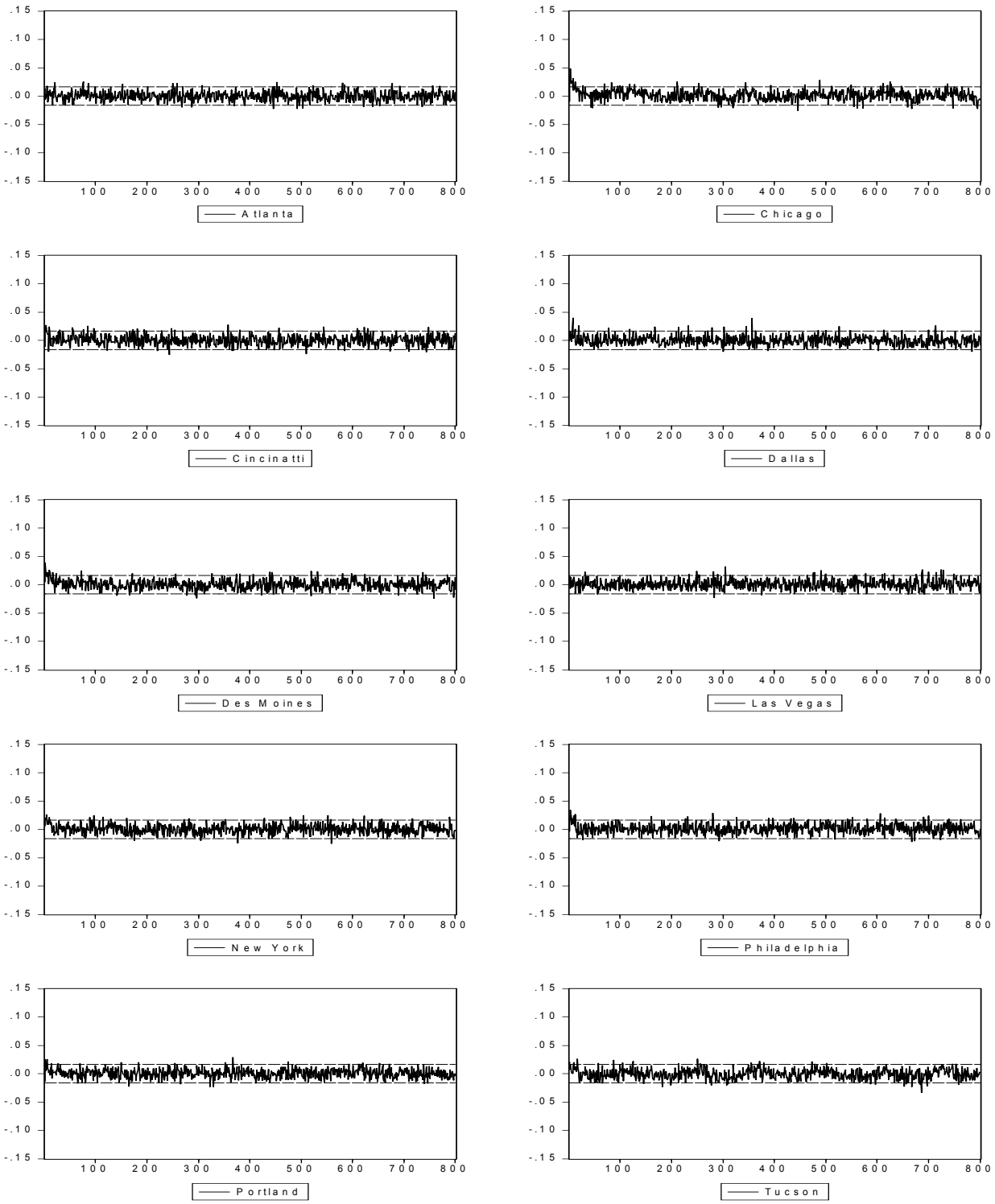
Notes: Each panel displays an empirical distribution of standardized residuals from our average temperature model, $\left(T_t - \hat{T}rend_t - \hat{S}easonal_t - \sum_{i=1}^{25} \hat{\rho}_i T_{t-i}\right) / \hat{\sigma}_t$.

Figure 11
Correlograms of Standardized Residuals
Daily Average Temperature



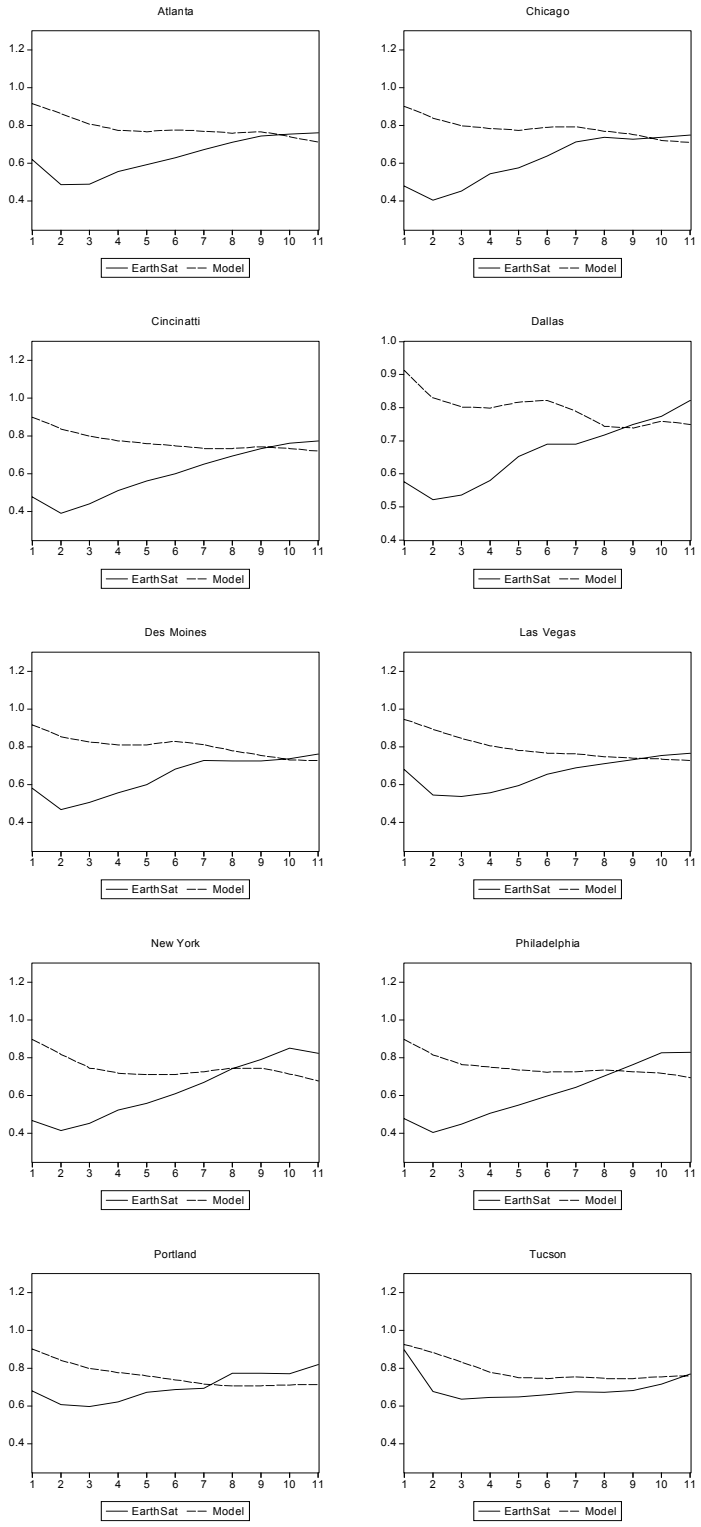
Notes: Each panel displays sample autocorrelations of the standardized residuals from our daily average temperature model, $\left(T_t - \hat{T}rend_t - \hat{S}easonal_t - \sum_{i=1}^{25} \hat{\rho}_i T_{t-i}\right) / \hat{\sigma}_t$, together with Bartlett's approximate ninety-five percent confidence intervals under the null hypothesis of white noise, as discussed in detail in the text.

Figure 12
Correlograms of Squared Standardized Residuals
Daily Average Temperature



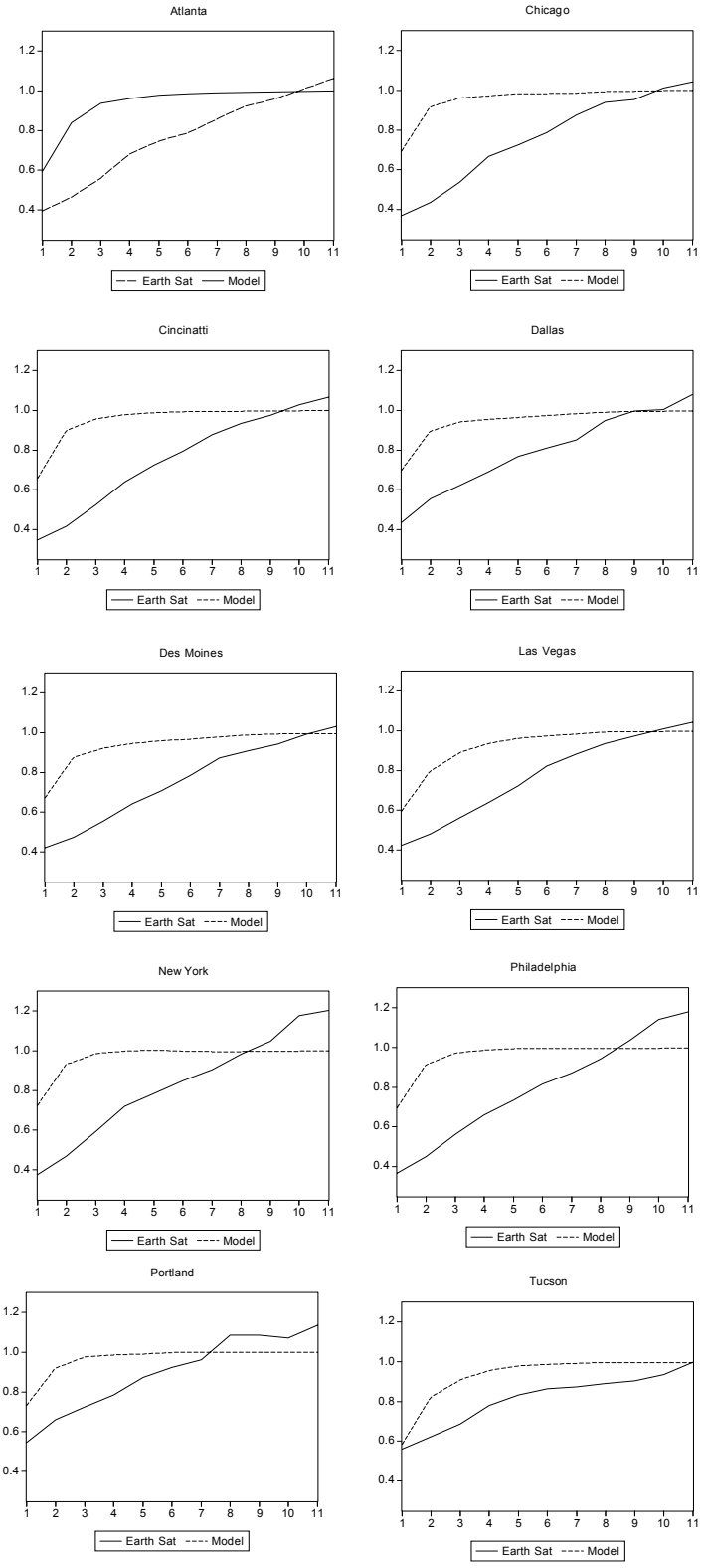
Notes: Each panel displays sample autocorrelations of the squared standardized residuals from our daily average temperature model, $\left(\left(T_t - \hat{T}rend_t - \hat{S}easonal_t - \sum_{i=1}^{25} \hat{\rho}_i T_{t-i}\right) / \hat{\sigma}_t\right)^2$, together with Bartlett's approximate ninety-five percent confidence intervals under the null hypothesis of white noise, as discussed in detail in the text.

Figure 13
Forecast Skill Relative to Persistence Forecast
Daily Average Temperature Point Forecasts



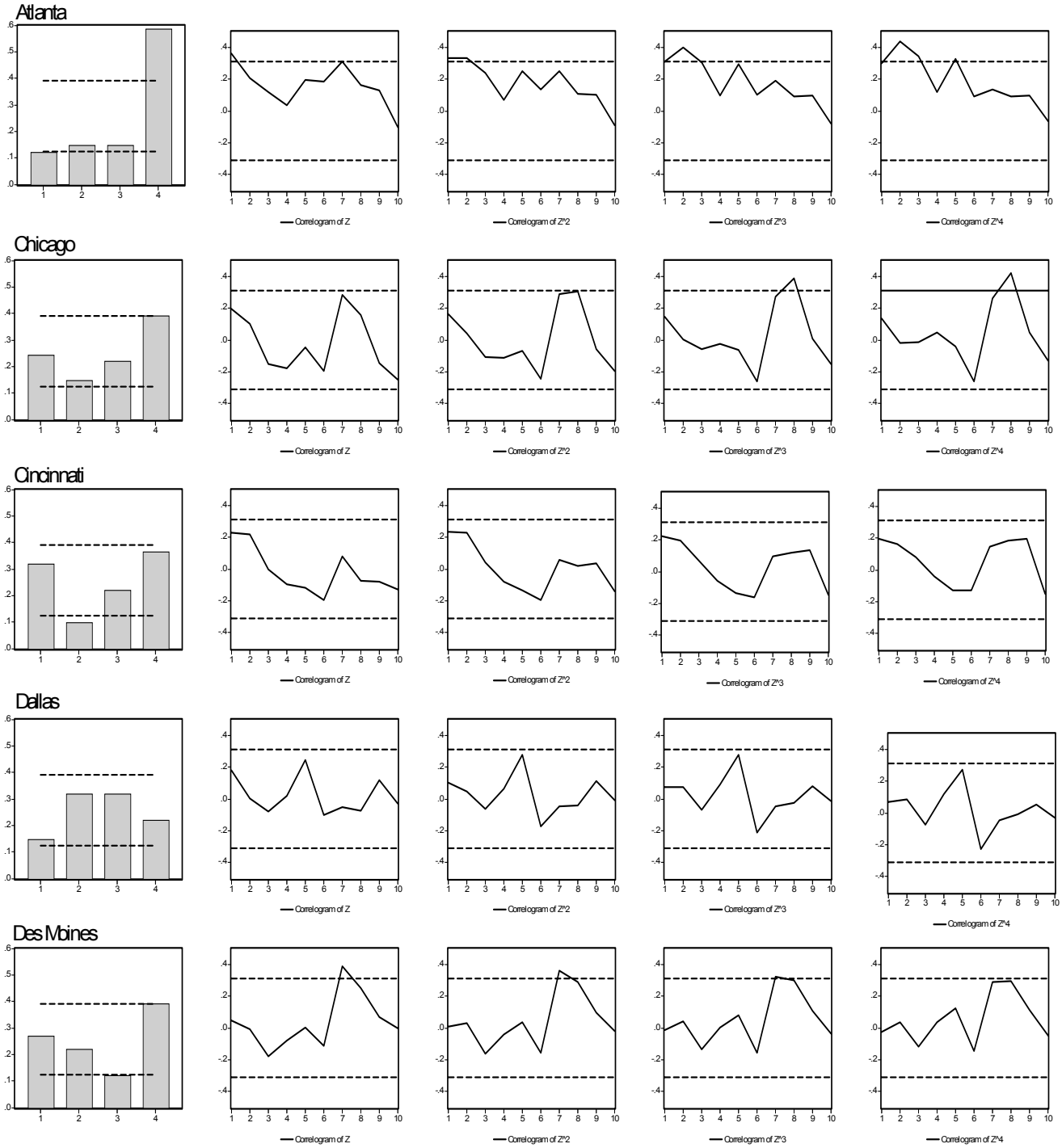
Notes: Each panel displays the ratio of a forecast's RMSPE to that of a persistence forecast, for 1-day-ahead through 11-day-ahead horizons. The solid line refers to the EarthSat forecast, and the dashed line refers to the autoregressive forecast. The forecast evaluation period is 10/11/99 - 10/22/01, as discussed in the text.

Figure 14
Forecast Skill Relative to Climatological Forecast
Daily Average Temperature Point Forecasts



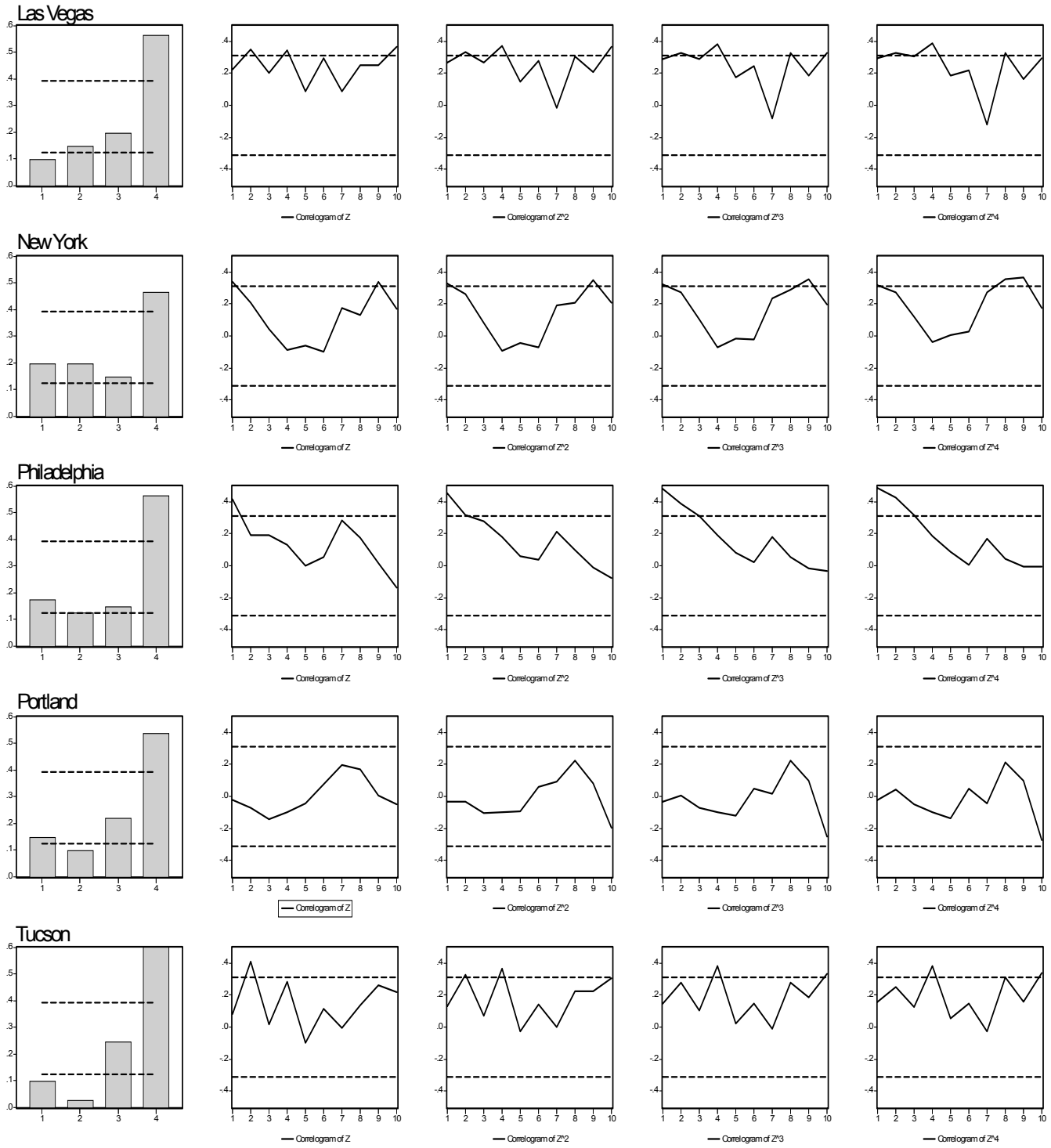
Notes: Each panel displays the ratio of a forecast's RMSPE to that of a climatological forecast, for 1-day-ahead through 11-day-ahead horizons. The solid line refers to the EarthSat forecast, and the dashed line refers to the autoregressive forecast. The forecast evaluation period is 10/11/99 - 10/22/01, as discussed in the text.

Figure 15
z-Statistics, Distributions and Dynamics
Daily Average Temperature Distributional Forecasts



Notes: Each row displays a histogram for *z*, as well as correlograms for the first four powers of *z*, where *z* is the probability integral transform of realized November-March HDDs, 1960-2001. Dashed lines indicate approximate ninety-five percent confidence intervals in the *iid U(0,1)* case corresponding to correct conditional calibration. See text for details.

Figure 15 (Continued)
z-Statistics, Distributions and Dynamics
Daily Average Temperature Distributional Forecasts



Notes: Each row displays a histogram for z, as well as correlograms for the first four powers of z, where z is the probability integral transform of realized November-March HDDs, 1960-2001. Dashed lines indicate approximate ninety-five percent confidence intervals in the *iid U(0,1)* case corresponding to correct conditional calibration. See text for details.