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 01-031

"Weather Forecasting for Weather Derivatives"

by

Francis X. Diebold and Sean D. Campbell

http://ssrn.com/abstract_id=284950

Weather Forecasting for Weather Derivatives

Francis X. Diebold
University of Pennsylvania
and NBER

Sean Campbell
University of Pennsylvania

December 2000
This revision/print: August 20, 2001

Abstract: Weather derivatives are a fascinating new type of Arrow-Debreu security, making pre-specified payouts if pre-specified weather events occur, and the market for such derivatives has grown rapidly. 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. Curiously, however, 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 be best for forecasting six hours ahead, it is not obvious that it is best for the longer horizons relevant for weather derivatives, such as six days, six weeks, or six months. In particular, good forecasting does not necessarily require a structural model. In this paper, then, we take a seemingly-naive nonstructural time-series approach to modeling and forecasting daily average temperature in ten U.S. cities, and we inquire systematically as to whether it proves useful. The answer is, perhaps surprisingly, "yes." Time series modeling reveals a wealth of information about both conditional mean dynamics and the conditional variance dynamics of average daily temperature, some of which seems not to have been noticed previously, and it provides similarly sharp insights into both the distributions of weather and the distributions of weather surprises, and the key differences between them. The success of time-series modeling in capturing conditional mean dynamics translates into successful point forecasting, a fact which, together with the success of time-series modeling in identifying and characterizing the distributions of weather surprises, translates as well into successful density forecasting.

Key Words: Risk management; Hedging; Insurance; Seasonality; Average temperature; Financial derivatives

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 René Garcia, Joseph Hrgovic, Vince Kaminski, Paul Kleindorfer, Howard Kunreuther, Don McIsaac, Nour Meddahi, Matt Pritsker, Claudio Riberio and Yihong Xia, as well as seminar participants at Enron and Wharton. 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. By 2000 the market had grown to $5 billion, with better liquidity. Outlets like *Energy and Power Risk Management* and the *Weather Risk* supplement 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 not unrelated). 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, as weather is by its nature a very location-specific and non-standardized commodity, unlike, say, as 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 long precipitation 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

---

[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.

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-removal costs, consumers who want fixed heating and air conditioning bills, firms seeking to avoid financial writedowns due to weather-driven poor performance, etc. The mere fact that such agents face weather fluctuations, however, does not ensure a large hedging demand, because even very large weather fluctuations create little weather risk if 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? Such questions are intimately related to 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. Leading players include such firms as Enron (the Houston-based giant), Koch Industries (a Wichita, Kansas, privately-held commodities giant), Utilicorp United (a Kansas City, Missouri utility) and Aquila Risk Management (its Omaha, Nebraska subsidiary), Natsource and Prebon Energy. But how should weather derivatives be priced? It seems clear that standard approaches to arbitrage-free pricing (e.g., Black-Scholes) are irrelevant 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 an approximate 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.[3] As Davis (2001) 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.

---

[3] In the special case of risk neutrality (e.g., a firm), it appears that no utility function is needed. What *is* needed is an estimate of the joint distribution of future weather and future revenues. From this one should be able to derive the firm's demand curve for the derivative.

[4] See also Cao and Wei (1999).

This again brings up the crucial issue of how to construct good weather forecasts, potentially at horizons much longer than those commonly adopted 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 overviews in Katz and Murphy, 1997) has a structural "atmospheric science" feel, and although such an approach may be best for forecasting six hours ahead, it is not at all obvious that it is best for the longer horizons relevant for weather derivatives, such as six days, six weeks, or six months. In particular, forecasting does not require a structural model, and in the last thirty years statisticians and econometricians have made great strides in the nonstructural (reduced-form) 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 (density forecasting). In this paper, then, we take a time-series approach to weather modeling and forecasting, systematically asking whether the seemingly-naive nonstructural approach proves useful.[6] The answer is indeed surprising. Time series modeling reveals a wealth of information about both conditional mean dynamics and the conditional variance dynamics of average daily temperature, some of which seems not to have been noticed previously, and it provides similarly sharp insights into both the distributions of weather and the distributions of weather surprises, and the key differences between them. Last, and not at all least, the success of time-series modeling in capturing conditional mean dynamics translates into successful point forecasting, a fact which, together with the success of time-series modeling in identifying and characterizing the distributions of weather surprises, translates as well into successful density forecasting.

We proceed as follows. In Section 2 we discuss our data, and our decision to focus on modeling and forecasting average daily temperature. In section 3 we report the results of time-series modeling of

---

[5] See Diebold (2001).

[6] We are not the first to adopt a time-series approach, although the literature is remarkably sparse. The rainfall analyses of Harvey (1989) and Hyndman and Grunwald (2000) suggest its value, for example, but curiously, they have not been widely appreciated. 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.

average daily temperature, and in section 4 we report the results of point forecasting and density forecasting competitions. 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 10/18/00, resulting in 14,902 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, minimum and average temperatures, together with heating degree days and cooling degree days, 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.

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

Modeling Daily Average Temperature

We choose to model and forecast daily average temperature, which is widely reported and followed. Moreover, the heating degree days (HDDs) and cooling degree days (CDDs) on which weather derivatives are commonly written are trivial transformations of average temperature, as shown in Table 2, so modeling and forecasting average temperature is effectively the same as modeling and forecasting HDDs or CDDs.

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, HDDs, CDDs, etc., given that one can simply model the underlying high-frequency temperature, from which all 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, tailored to the object of interest, 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.

Although we perform our average temperature modeling on a daily basis, we will use our estimated daily models to forecast at significantly longer horizons. Even if forecastability drops quickly with horizon – an empirical issue that we shall explore – there could still be important gains from exploiting it, especially when the time to maturity is short (e.g., in HDD forecasting, forecast object is typically the *sum* of HDDs across the horizon).

3. Time-Series Weather Modeling

Before proceeding to detailed modeling and forecasting results, it is useful to get an overall view of the daily average temperature data. In Figure 1 we plot the average temperature series for the last five

---

[8] For an introduction to forecasting under the relevant loss function, see Diebold (2001).

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. The Midwestern cities of Chicago and Des Moines have the most volatile temperatures, while the western cities of Portland, OR and Tucson are the least volatile.

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 important given the rather large size of our dataset (roughly 15,000 daily observations for each of ten cities) and the thousands of models over which we subsequently search.

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 one suspects may be relevant but is probably 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, which again proves crucial due to the large number of models examined.[10]

[9] We could of course have plotted the average temperature series over the entire sample, but doing so results in a less informative graph.

[10] ARMA models would in principle enable greater parsimony, but only at the (very large) expense of requiring numerical optimization for parameter estimation. In the context of the empirical work of this paper, we would much prefer to be in the business of estimating, say, AR(10) models than

Assembling the various pieces, we estimate the following model for average temperature in each of our ten cities between 1/1/1960 and 10/1/1999:[11]

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

where

$$Trend_t = \beta_0 + \beta_1 t \qquad (1a)$$

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

$$\varepsilon_t \sim (0, 1), \qquad (1c)$$

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, including up to thirty autoregressive lags (L=30) and five Fourier sine and cosine terms (P=5), chosen using the Akaike and Schwarz information criteria (AIC and SIC). AIC tends to choose models with substantially greater L than does SIC, whereas the AIC and SIC choose similar values of P.

Now let us discuss the estimation results.[13] 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 (with order selected by AIC) against the estimated dummy variable seasonal pattern. 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 average daily temperature. In most cases, the trend is much larger than the increase in average global

---

ARMA(2,2) models.

[11] Note that we have reserved approximately one year of data – not used in model selection or estimation – for subsequent out-of-sample forecast analysis.

[12] We removed February 29 from each leap year in our sample to maintain 365 days per year.

[13] For the most part, we limit our discussion to AIC-selected models; the results for SIC-selected models are qualitatively very similar. We present complete estimation results in Appendix A.

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]

Average daily temperature also displays strong cyclical persistence. The estimated autoregressive coefficients in AIC-selected models 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 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 next roots are always complex, with modulus decaying slowly, as shown in Figure 4. The SIC, in contrast, always selects three autoregressive lags, corresponding to a dominant real root and a pair of complex conjugate roots, as also shown in Figure 4.

In Figure 5 we show fitted values and residuals over the last five years of the estimation sample for the models selected by AIC. The fit is typically very good, with R^2 above 90%. Figure 5 also provides a first glimpse of an important and apparently little-known phenomenon: pronounced and persistent time-series variation in the variance of the residuals. Put differently, 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 selected by AIC. Four features stand out. First, the residuals are much less variable than the average temperature; that is, weather surprises are much less variable than the weather itself.[16] Second, the spreads of the distributions vary noticeably across cities; the standard deviations range from a low of 3.5 degrees for Tucson to a high of

---

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

[15] It seems that such seasonal heteroskedasticity was first noted by Roll (1984). In this paper we go farther, developing quantitative models and forecasts that incorporate the heteroskedasticity.

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

6.2 degrees for Des Moines. 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 lags. The residual autocorrelations are negligible and appear consistent with white noise, indicating that we have modeled linear conditional mean dynamics 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]

We therefore move to the model,[18]

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

where

$$Trend_t = \beta_0 + \beta_1 t \tag{2a}$$

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

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

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

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). We estimated models including up to thirty autoregressive lags (*L=30*), five Fourier sine and cosine terms in both the seasonal and the variance specification

[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.

[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 from model (2) were conditionally heteroskedastic.

(P=Q=5), and five lagged squared residuals in the variance equation (R=5), for a total of 6,696 models.

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 seemed evident 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, which often arises 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.[19] 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 squared residuals from equation (2) on constant, Fourier and lagged squared residual terms, and we use the square root of the inverse fitted values, σ_t^{-1}, as weights in a weighted least squares re-estimation of equation (2).

We choose the lag lengths L, P, Q and R using the AIC and SIC. In Table 4 we report the models selected by each criterion for each city. Generally, as expected, AIC chooses models with more lagged dependent variables (L). The largest model chosen contains L=29 lags (Dallas), whereas the largest model selected by SIC contains only L=3 lags (Atlanta, Tucson). The AIC and SIC agree more closely in selecting P, Q and R; the difference in the number of lagged squared residuals (R) chosen by the AIC and SIC, for example, is typically only two or three.

In Figure 9 we plot the estimated residual conditional standard deviation from 1993 through 1998. The basic pattern is one of seasonal volatility variation, with short-lived upward-brushing ARCH effects.[20] For each city, seasonal volatility appears highest during the winter months. Some cities display a great deal of seasonal volatility variation – the conditional standard deviations in Atlanta, Cincinnati

[19] Full asymptotic efficiency is known to be achieved if ε_t is Gaussian; see Engle (1982).

[20] In sharp contrast to the standard situation in financial econometrics (see, for example, Andersen, Bollerslev and Diebold, 2001), 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. This appears to be the case, for example, in the results of Torro, Meneu and Valor (2001).

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 seasonal volatility of weather surprise volatility increases with its average level. 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, although modeling conditional heteroskedasticity reduces residual excess kurtosis, it does not eliminate it. As expected, moreover, each distribution still displays negative skewness.

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), so it is not surprising that there is none in the 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 an apparently good time-series model for daily average temperature, we now proceed to swallow hard and march onward, using it for weather forecasting. We explore the forecasting accuracy of our models at horizons ranging from one day through thirty days, examining both point and density forecasts, in-sample and out-of-sample.

<u>Point Forecasting</u>

We assess the root mean squared error (RMSE) accuracy of our average daily temperature forecasts (for both AIC- and SIC- selected models) relative to three benchmark competitors. The first benchmark is a simple deterministic seasonal model that allows for daily effects via day-of-year dummy variables, in keeping with the ubiquitous meteorological use of "daily averages" as benchmarks. The second is a "seasonal plus trend" model, containing daily dummies and a linear trend. The third is a random walk, which of course produces "no-change" forecasts. We estimate all models using the two-step weighted least squares approach described earlier.

We track forecast accuracy at 1-day, 7-day and 30-day horizons. Because weather derivatives are often written in terms of the cumulative sum of HDDs and CDDs relative to a "strike" level, we consider the cumulative sum of daily average temperature over an h-day horizon when evaluating h-step-ahead forecasts. In the case of the 1-day horizon we simply consider the forecast errors, $\hat{\varepsilon}_{t+1|t} = T_{t+1} - \hat{T}_{t+1|t}$. In the case of multi-step-ahead forecasts we proceed as follows. For each day *j* over the horizon of *h=7* or *h=30* days, we compute the *j*-step-ahead average temperature forecast $\hat{T}_{t+j|t}$ and the corresponding *j*-step-ahead forecast error $\hat{\varepsilon}_{t+j|t} = T_{t+j} - \hat{T}_{t+j|t}$, *j= 1, ..., h*. Then we sum the forecast errors over the horizon to construct the cumulative *h*-day forecast error $\hat{\varepsilon}_t^h = \sum_{j=1}^{h} (T_{t+j} - \hat{T}_{t+j|t})$. We do all of our multi-step forecasting in a non-overlapping fashion, resulting in a set of non-overlapping multi-step forecast errors for each city, horizon and model.

In Table 5 and Figures 13 and 14 we display Theil's U-statistics for in- and out-of-sample forecast errors. The U-Statistic is simply the ratio of a given forecast's RMSE to the random walk forecast's RMSE. Consider first the statistics for in-sample forecast errors. By "in-sample" we refer to the fact that $\hat{T}_{t+j|t}$ is based on parameter estimates that use all of the data from 1/1/1960 and 10/1/1999, and the RMSE calculations are done on that same sample. The forecasting performance of the Seasonal and Seasonal + Trend models starts out much worse than the random walk at the 1-day horizon, but the situation is reversed as we move to 7- and 30-day horizons. The forecasting performance of the our AIC-selected and SIC-selected is almost always better than that of all other models, and it improves with horizon relative to the random walk and worsens with horizon relative to the Seasonal and Seasonal + Trend models. Performance of the AIC- and SIC-selected models is about the same. The results are very intuitive: the performance of the random walk forecast worsens with horizon, because persistent but mean-reverting daily average temperature (as we documented earlier) implies that a no-change forecast will perform well at short horizons (due to the high persistence) but less well at longer horizons (due to the failure of the no-change forecast to capture mean reversion).[21]

Now consider the U-statistics for out-of-sample forecast errors. In addition to in-sample forecasting performance, it is also of interest to evaluate out-of-sample forecasting performance. The benefit of doing so is guarding against data mining that may spuriously enhance in-sample performance; the cost is a small number of out-of-sample observations. By "out-of-sample" evaluation we refer to the fact that although $\hat{T}_{t+j|t}$ is based on parameter estimates that use data from 1/1/1960 through 10/1/1999, but the RMSE calculations are done on the *later* sample, from 10/2/1999 through 10/18/2000. The in-sample results continue to hold: our models almost always produce forecasts that are superior to the benchmarks, and their superiority relative to the random walk increases with horizon, while their superiority relative to the Seasonal and Seasonal + Trend models decreases with horizon. The good

---

[21] Seasonal fluctuations, which the random walk also misses, are also more relevant at longer horizons.

forecasting performance at all horizons is made clear in Figures 15-17, in which we show scatterplots of realizations against forecasts for all cities and all horizons.

Density Forecasting

We now assess the adequacy of 1-day-ahead density forecasts produced by our forecasting model.[22] Effectively we form the density forecasts by simulation. We simulate a 1-day-ahead realization by taking the 1-day-ahead point forecast (the construction of which was described earlier) and adding a draw from the historical distribution of the 1-day-ahead forecast errors. We repeat this many times and then we estimate the density of the 1-step-ahead realization, which we use as our density forecast. We judge density forecast adequacy in terms of conditional calibration and use the methods for checking correct conditional calibration developed by Diebold, Gunther and Tay (1998) and Diebold, Hahn and Tay (1999). Those methods involve computing the probability integral transform z_t of the data, and checking whether it is iid U(0,1). To that end, we examine the properties of z_t for Philadelphia (our home city) in Figure 18, which contains a histogram of z_t and correlograms of the first four powers of z_t, for each of the five competing models outlined above.[23] The figure confirms that our AIC- and SIC-selected models are correctly conditionally calibrated; the histograms look uniform and the correlograms look flat. The situation is reversed for the benchmark models; in particular, the large sample autocorrelations at the first few lags for each power of z_t indicate poor conditional calibration.

5. Concluding Remarks and Directions for Future Research

Motivated by the centrality of weather modeling and forecasting beneath both the supply and demand of weather derivatives, we have provided a succinct stochastic characterization of the distribution and dynamics of average daily temperature, and temperature "surprises," for ten major U.S. cities. Although we have learned a great deal and produced weather models that are both simple and powerful, it is clear that our analysis is at best a small first step. Many issues remain unexplored. Here we list a few of those that we find most intriguing, in no particular order.

Other Weather Variables

We have of course examined only daily average temperature. It will be of interest also to model the daily maximum and minimum series using the relevant extreme value theory. The underlying high-frequency temperature data are highly serially correlated, which significantly complicates the standard

---

[22] Longer horizons are infeasible, given that a large number of observations are needed for successful application of the conditional calibration tests.

[23] Parallel figures for all of the remaining cities appear in Appendix B.

extreme value theory. It will similarly be of interest to examine other weather variables such as precipitation, etc.

Multivariate Analysis and Cross-Hedging

We have already briefly touched upon multivariate analysis in the context of jointly modeling daily maximum and minimum temperatures for a given city. Cross-city correlations may be crucially important as well, 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 analysis reported in the present paper is vector autoregressive modeling 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. For example, Bodoukh, Sjen and Whitelaw (2001) study nonlinearities in the relationship between prices and the weather, induced via path dependence, noting that 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.

Statistical Modeling Issues

A wealth of modeling issues merit additional investigation, including 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, long memory dynamics, structural stability assessment, and allowing for gradual parameter evolution.

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 use for forecasting at all horizons. Alternatively, we could use a daily model for 1-day-ahead forecasts, a weekly model for 7-day-ahead forecasts, and so on, fitting customized models at each horizon.

Additional Forecast Comparisons

We would very much like to compare the performance of our out-of-sample forecasts not only to that of the benchmark competitors examined here, but also to that of a more sophisticated competitor, in particular, a real-time "industry standard" forecast, such as those produced by the National Weather Service. Thus far it has proved challenging, however, to produce an acceptable comparison with identical forecast variables, forecast horizons and information sets.

World-Wide Weather Derivatives Potential

We have assessed some of the factors that drive the demand for derivatives in the U.S., but it will be interesting to quantify world-wide potential for weather derivatives. Hedging demand, for example, might be higher in Asia due to greater weather noise associated with the presence of monsoons.

Optimal Aggregation Levels

Related, we need to determine the optimal amount of spatial aggregation for successful longer-term weather forecasting. Is it counties, cities, metro areas, states, regions, countries, or world regions?

References

Andersen, T.G., Bollerslev, T. and Diebold, F.X. (2001), "Parametric and Nonparametric Volatility Measurement," in L.P. Hansen and Y. Ait-Sahalia (eds.), *Handbook of Financial Econometrics*. Amsterdam: North-Holland, in progress.

Cao, M. and Wei, J. (1999), "Pricing Weather Derivatives: An Equilibrium Approach," Manuscript, University of York and University of Toronto.

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

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.

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.

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

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.

Katz, R.W. and Murphy, A.H., Eds. (1997), *Economic Value of Weather and Climate Forecasts*. Cambridge: Cambridge University Press.

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

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.

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

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

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

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

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	Maximum Daily Temperature - rounded to the nearest integer
TMIN	Minimum Daily Temperature - rounded to the nearest integer
TAVG	(TMAX+TMIN)/2
HDD	max(0, 65-TAVG)
CDD	max(0, TAVG-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 above 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 30 year mean difference in the missing variable as well as the 30 year standard deviation in the difference.

Table 4
Model Selection Results

City	AIC				SIC			
	Conditional Mean		Conditional Variance		Conditional Mean		Conditional Variance	
	Seasonal (P)	AR (L)	Seasonal (Q)	AR (R)	Seasonal (P)	AR (L)	Seasonal (Q)	AR (R)
Atlanta	4	21	3	1	3	3	2	1
Chicago	5	7	4	5	2	3	1	3
Cincinnati	5	14	4	5	1	3	1	0
Dallas	5	29	5	3	2	3	2	1
Des Moines	3	10	4	5	2	3	1	2
Las Vegas	3	4	4	2	2	3	2	1
New York	3	7	3	5	1	3	1	1
Philadelphia	3	7	5	5	1	3	1	0
Portland, OR	4	4	2	3	2	3	0	1
Tucson	5	9	5	1	3	3	3	0

Notes: For each city, we selected P, L, Q and R in the following model using both the Akaike and Schwarz criteria:

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

where

$$Trend_t = \beta_0 + \beta_1 t$$

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

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

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

See text for details.

Table 5
In- and Out-of-Sample Point Forecast RMSE Comparisons

	Seasonal		Seasonal + Trend		Seasonal + Trend + Cycle (AIC)		Seasonal + Trend + Cycle (SIC)	
	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
1 Day Ahead								
Atlanta	1.41	1.49	1.39	1.46	0.90	0.89	0.90	0.89
Chicago	1.34	1.31	1.33	1.29	0.90	0.90	0.90	0.90
Cincinnati	1.29	1.35	1.29	1.34	0.90	0.89	0.90	0.89
Dallas	1.35	1.42	1.35	1.38	0.89	0.89	0.90	0.90
Des Moines	1.37	1.40	1.38	1.36	0.91	0.92	0.91	0.92
Las Vegas	1.59	1.82	1.57	1.63	0.93	0.93	0.93	0.93
New York	1.24	1.26	1.23	1.25	0.89	0.89	0.89	0.89
Philadelphia	1.40	1.28	1.39	1.26	0.92	0.91	0.92	0.91
Portland, OR	1.27	1.31	1.26	1.31	0.90	0.89	0.90	0.89
Tucson	1.48	1.67	1.46	1.54	0.91	0.91	0.91	0.91
7-Day-Ahead								
Atlanta	0.75	0.97	0.73	0.94	0.66	0.76	0.67	0.76
Chicago	0.72	0.76	0.72	0.73	0.66	0.61	0.66	0.63
Cincinnati	0.72	0.81	0.72	0.80	0.65	0.67	0.66	0.68
Dallas	0.68	0.85	0.68	0.81	0.63	0.67	0.63	0.71
Des Moines	0.72	0.93	0.73	0.88	0.66	0.71	0.67	0.74
Las Vegas	0.81	1.14	0.80	0.98	0.71	0.79	0.71	0.79
New York	0.67	0.87	0.67	0.85	0.62	0.74	0.62	0.75
Philadelphia	0.79	0.72	0.78	0.71	0.69	0.65	0.70	0.65
Portland, OR	0.71	0.83	0.69	0.83	0.64	0.73	0.65	0.75
Tucson	0.78	0.95	0.76	0.83	0.69	0.73	0.69	0.72
30-Day-Ahead								
Atlanta	0.40	0.35	0.38	0.27	0.36	0.27	0.37	0.25
Chicago	0.38	0.59	0.38	0.54	0.36	0.47	0.37	0.49
Cincinnati	0.37	0.37	0.36	0.35	0.36	0.32	0.37	0.36
Dallas	0.36	0.56	0.36	0.51	0.35	0.42	0.35	0.49
Des Moines	0.40	0.68	0.39	0.62	0.37	0.55	0.38	0.58
Las Vegas	0.38	0.78	0.37	0.57	0.36	0.49	0.36	0.46
New York	0.36	0.36	0.34	0.32	0.33	0.31	0.34	0.32
Philadelphia	0.40	0.30	0.39	0.27	0.38	0.29	0.38	0.29
Portland, OR	0.38	0.38	0.36	0.36	0.35	0.34	0.36	0.36
Tucson	0.38	0.50	0.35	0.35	0.34	0.30	0.35	0.31

Notes: We show Theil's U-Statistics, the ratio of a forecast's RMSE to that of a random walk.

Figure 1
Daily Average Temperature

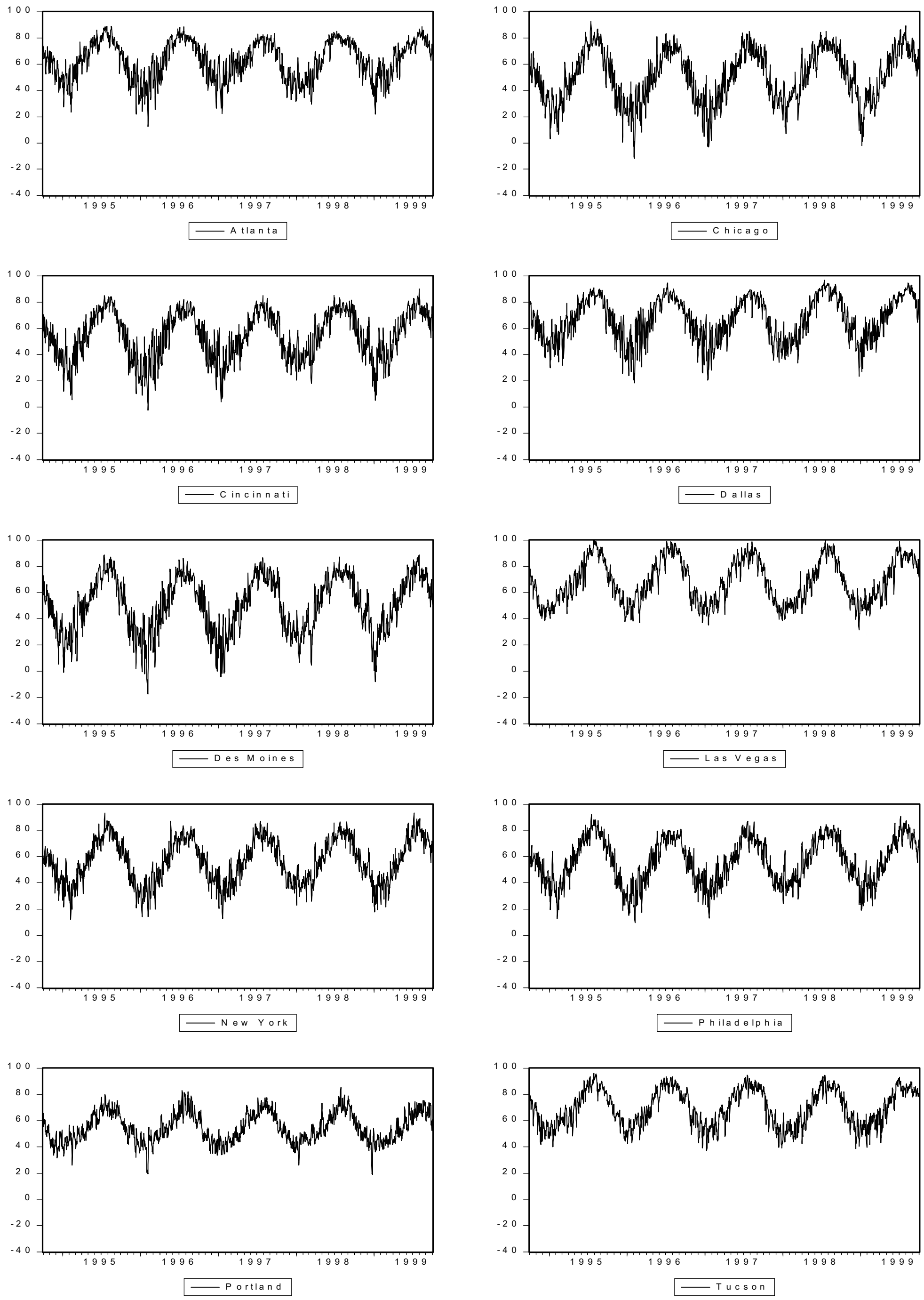


Figure 2
Estimated Unconditional Distribution of Daily Average Temperature

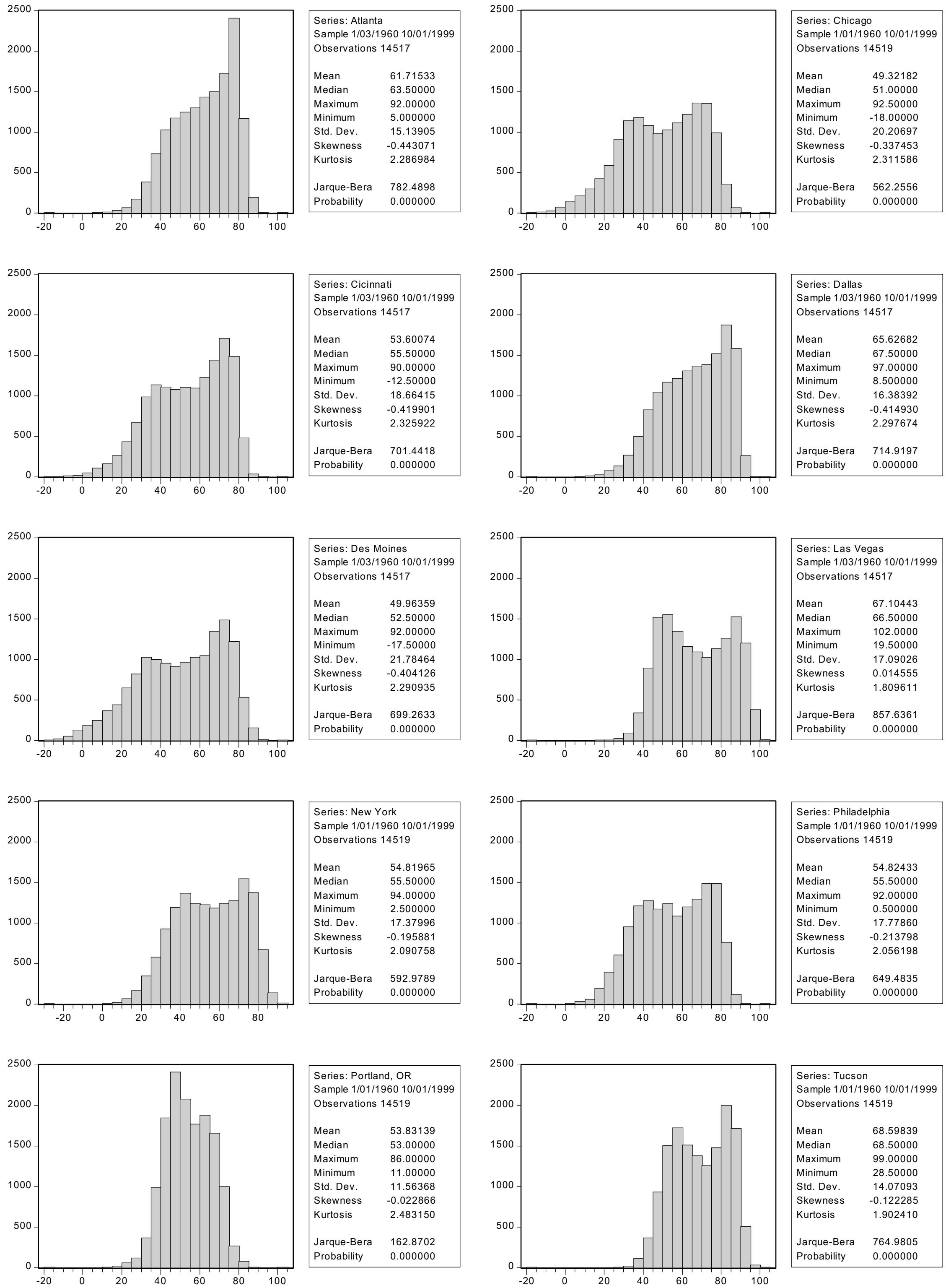


Figure 3
Daily Dummy Seasonal vs. Fourier Seasonal

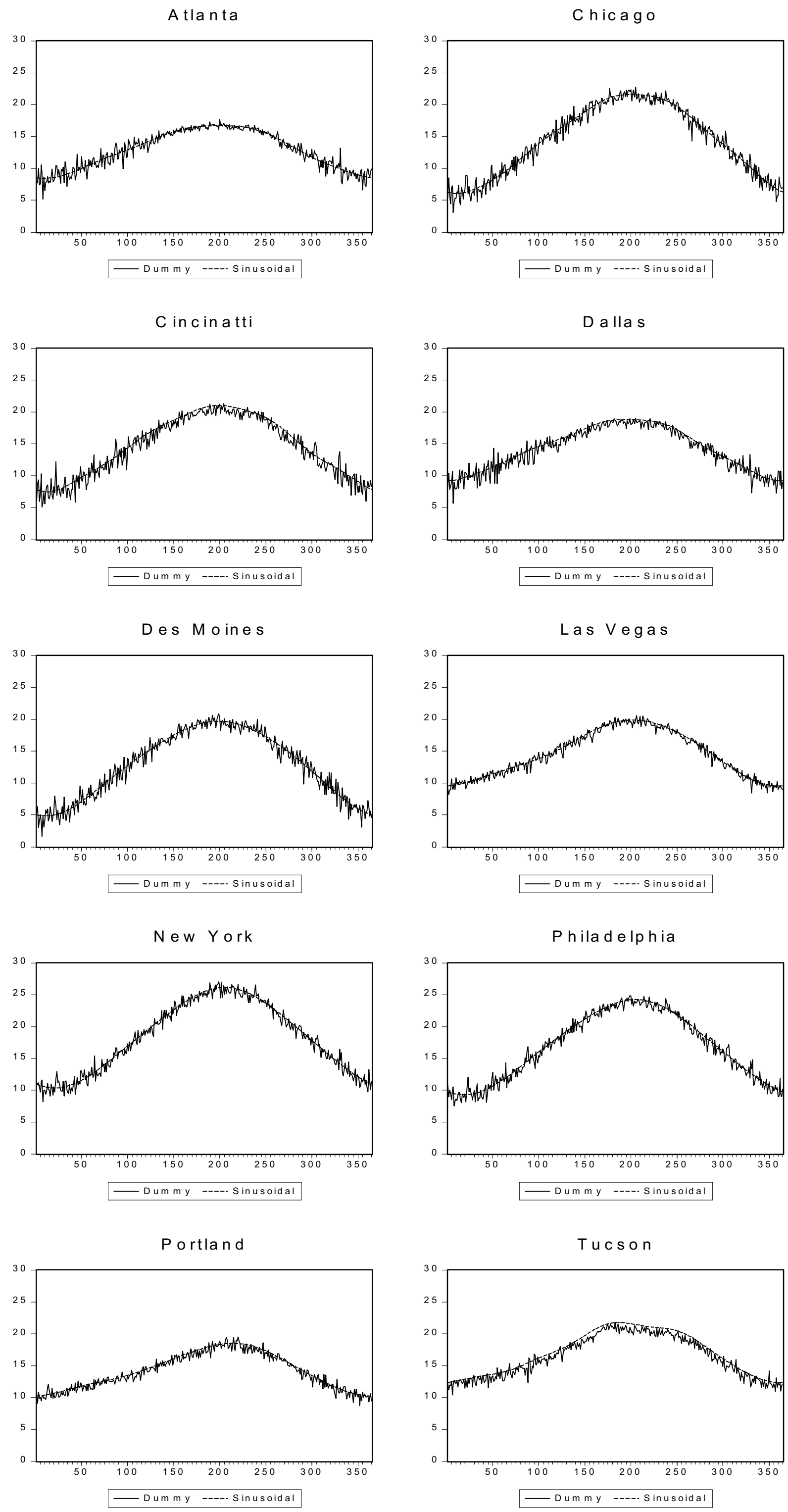
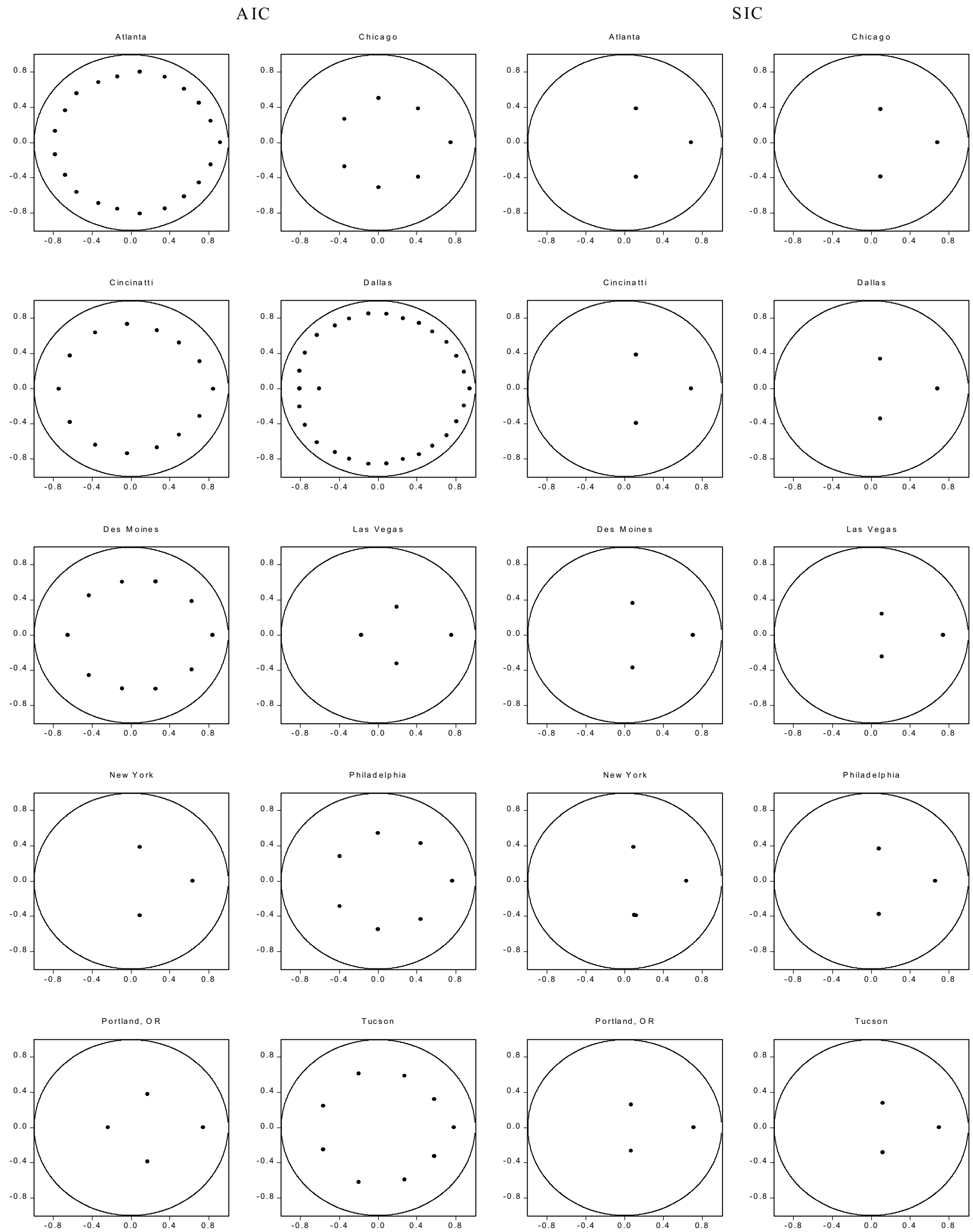


Figure 4
Reciprocal Roots of AIC- and SIC-Selected Autoregressive Lag Operator Polynomials



Notes: We plot the reciprocal roots of the autoregressive polynomial in the backshift operator, $1-\sum_{l=1}^{L} \rho_{t-l} B^{l}$, associated with the model (2), with lag order L chosen by AIC (left panels) and SIC (right panels), with the unit circle superimposed for visual reference. The real component of each inverse root appears on the horizontal axis, and the imaginary component appears on the vertical axis.

Figure 5
Daily Average Temperature
Fitted Values and Residuals

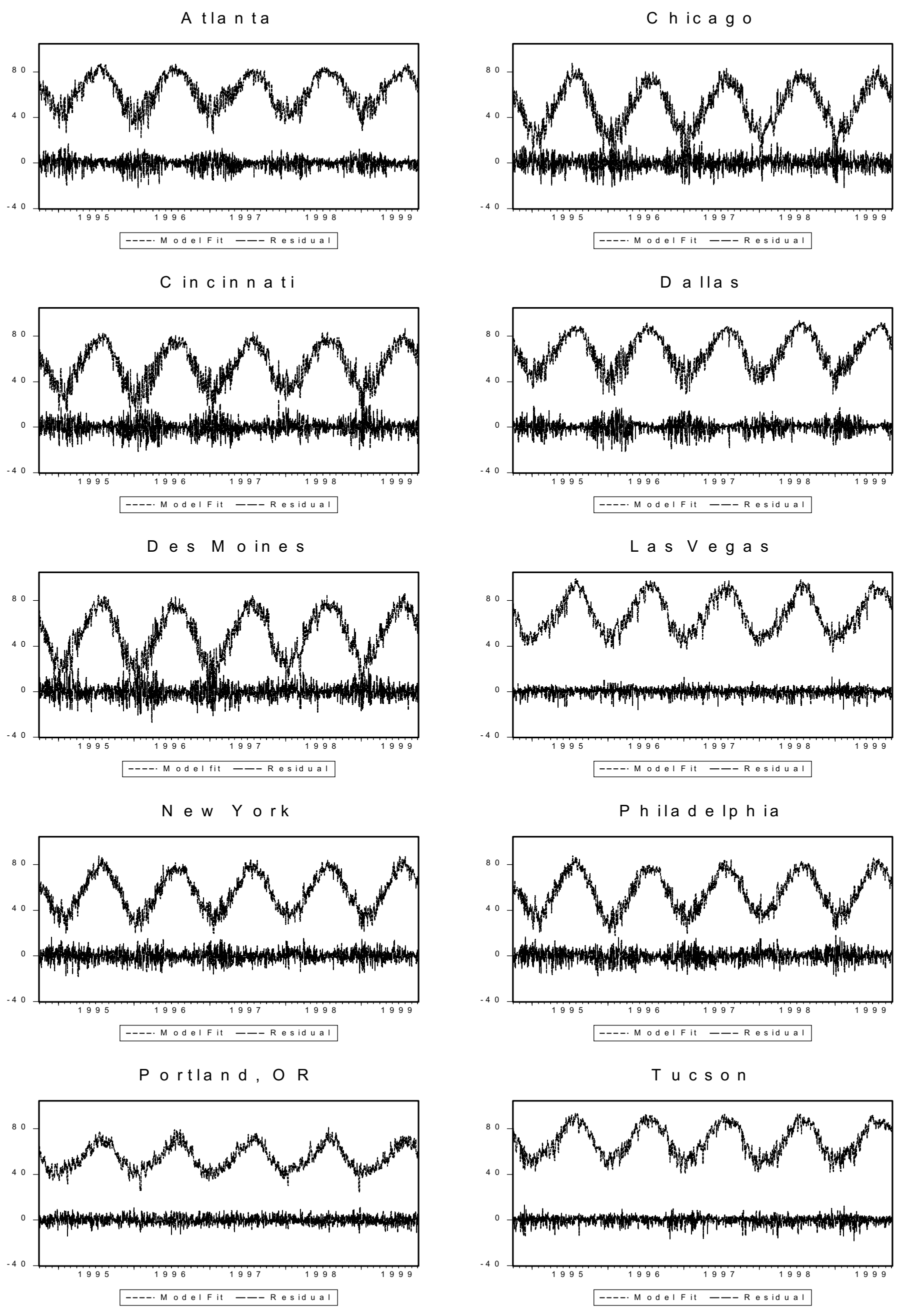


Figure 6
Estimated Unconditional Distributions of Residuals

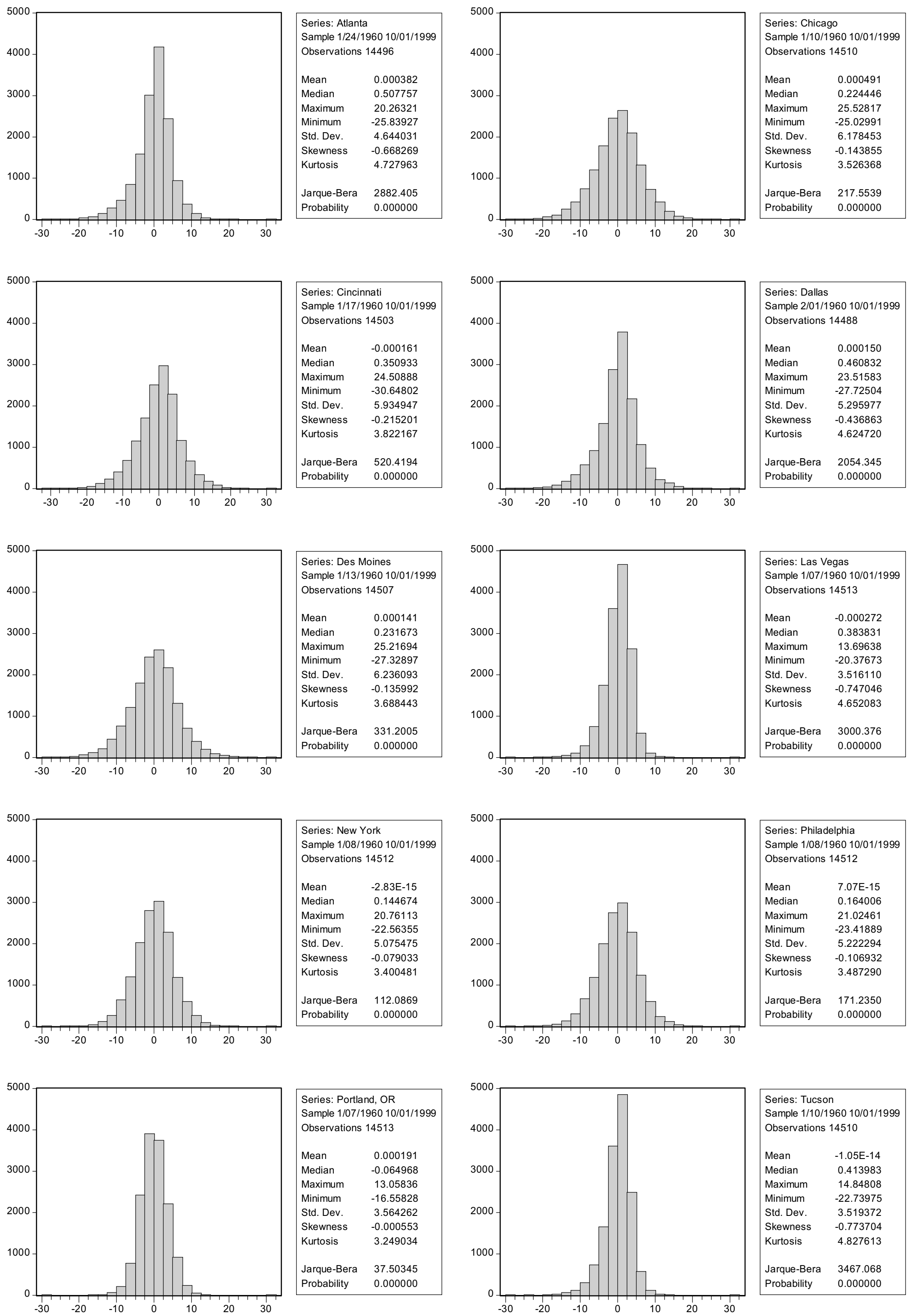


Figure 7
Correlograms of Residuals

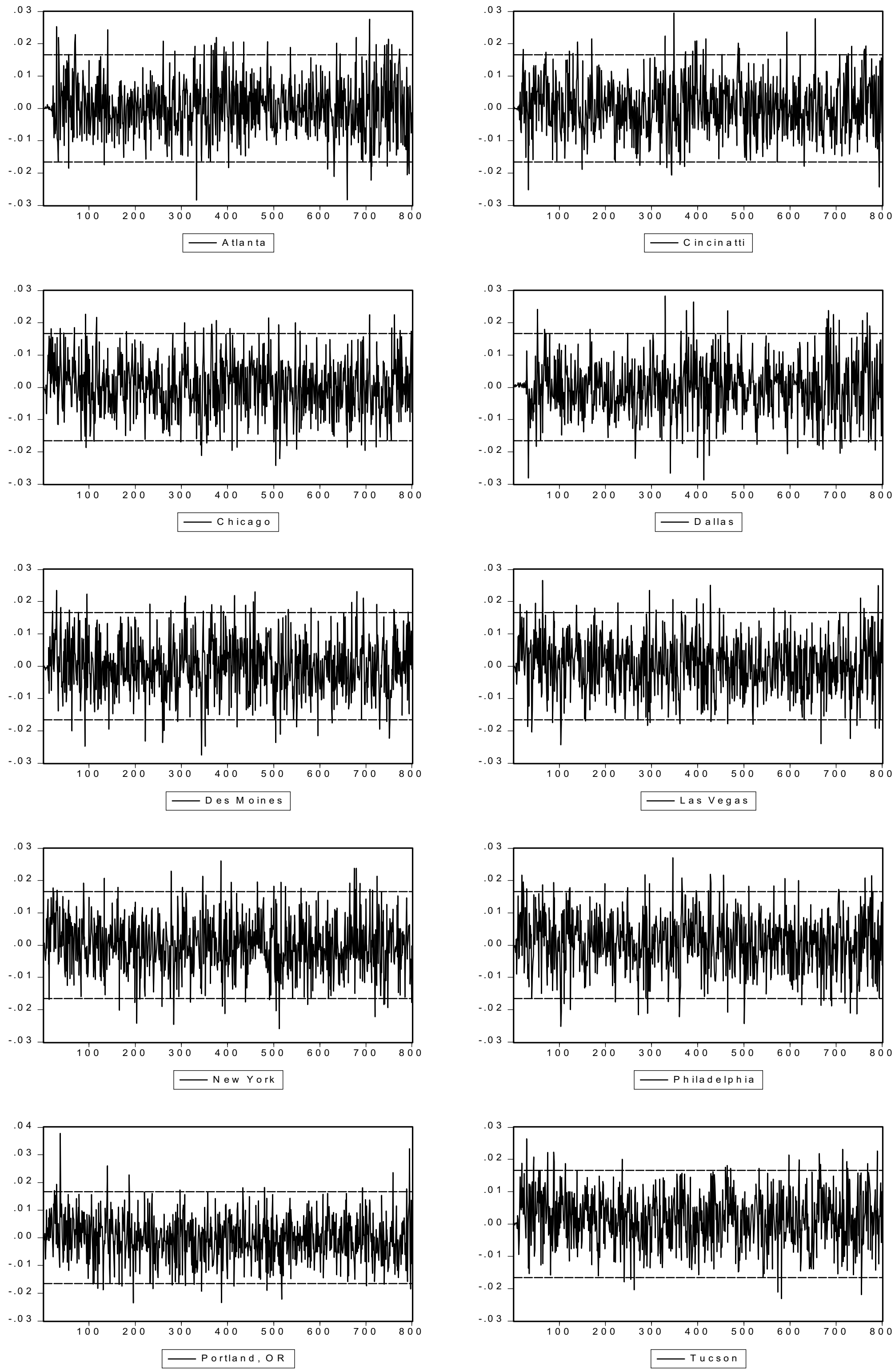


Figure 8
Correlograms of Squared Residuals

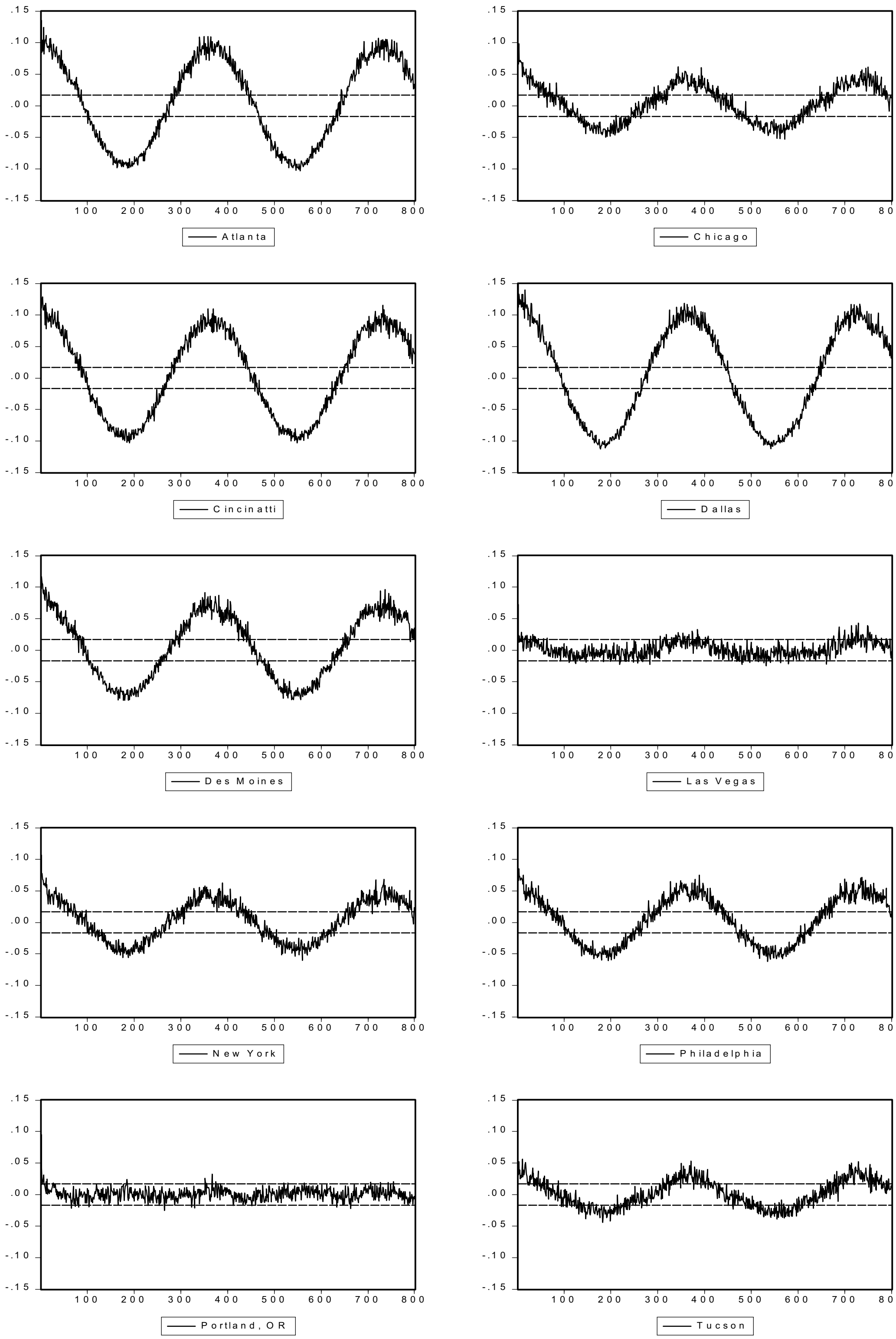


Figure 9
Estimated Conditional Standard Deviations of Residuals

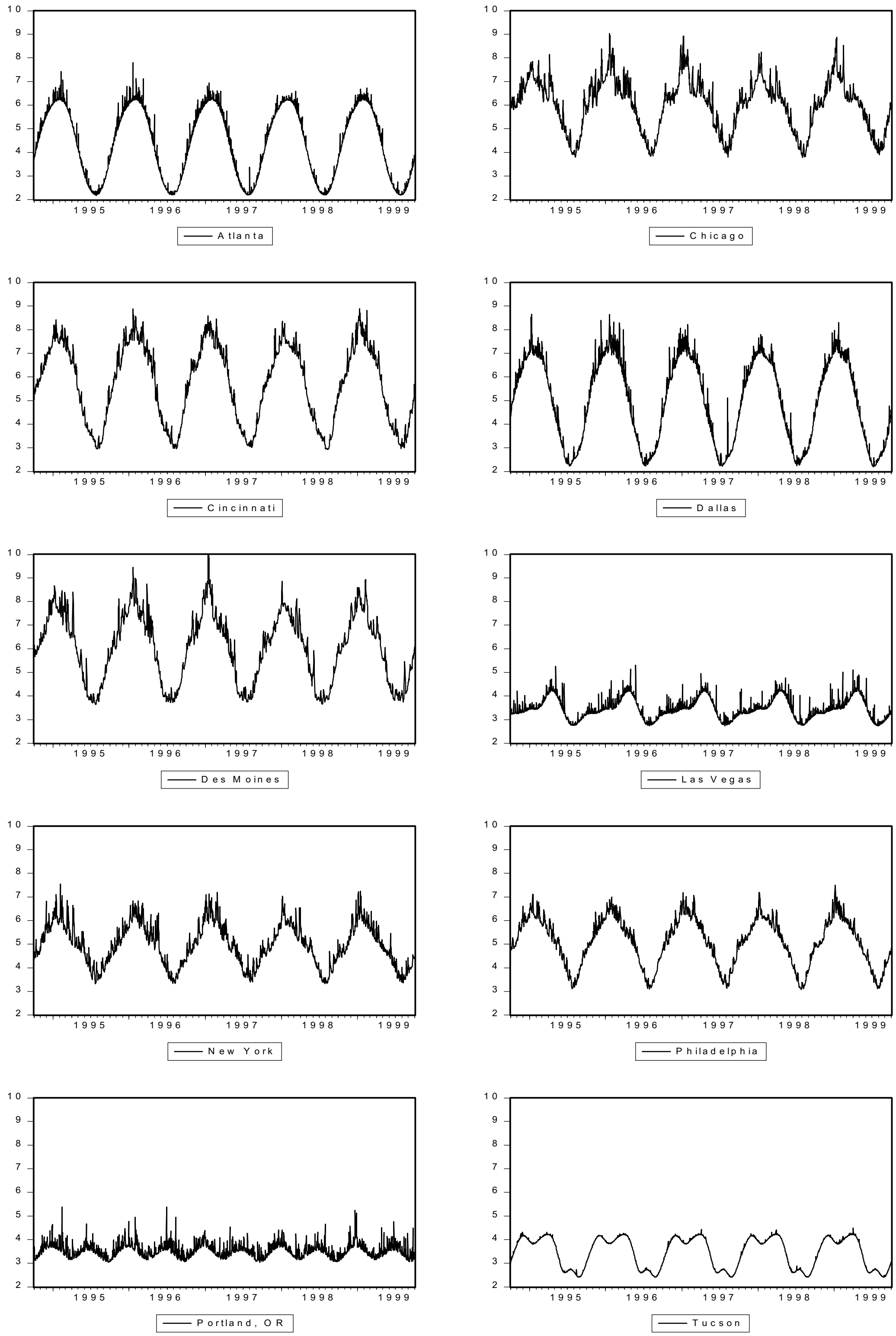


Figure 10
Estimated Unconditional Distributions of Standardized Residuals

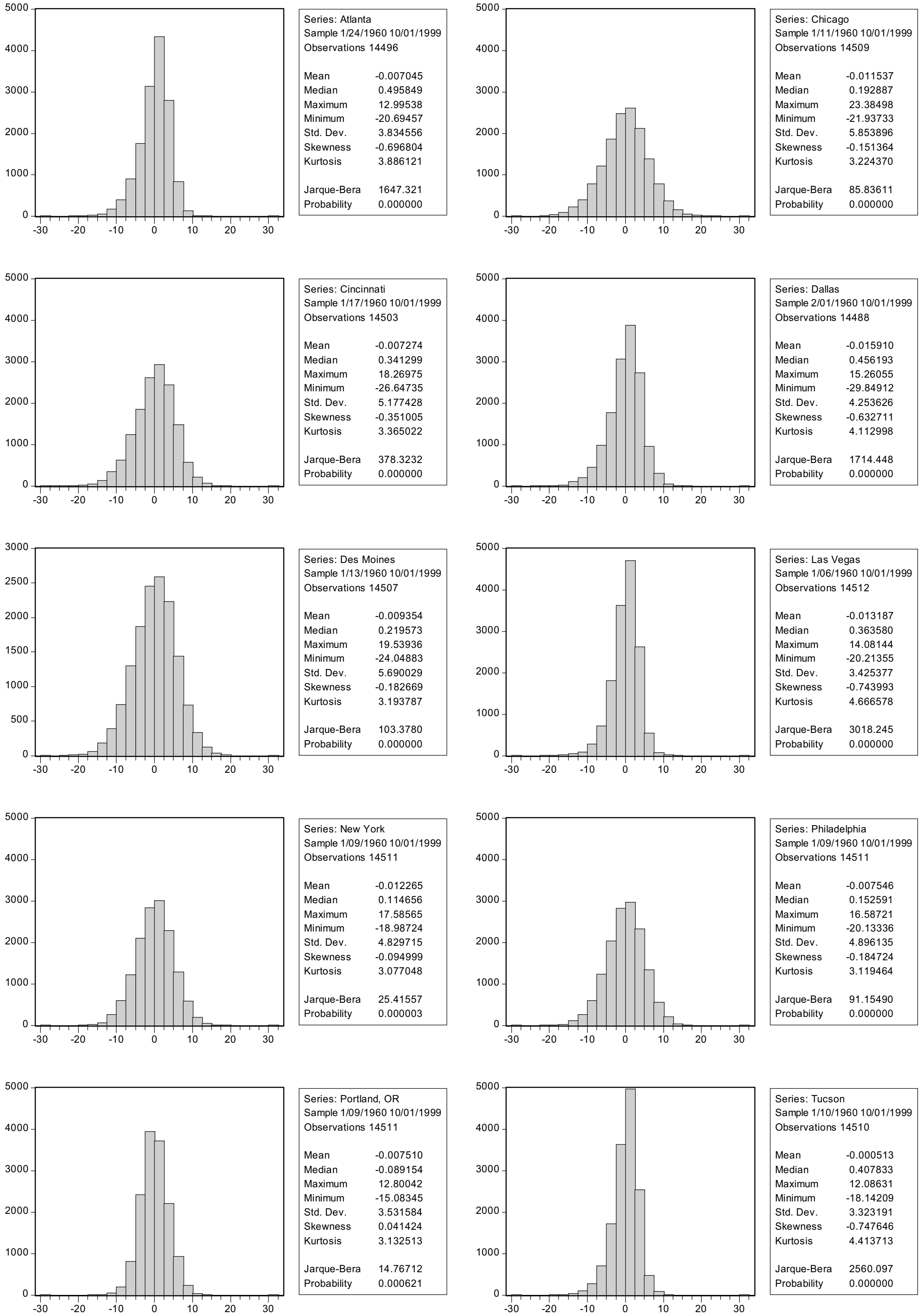


Figure 11
Correlograms of Standardized Residuals

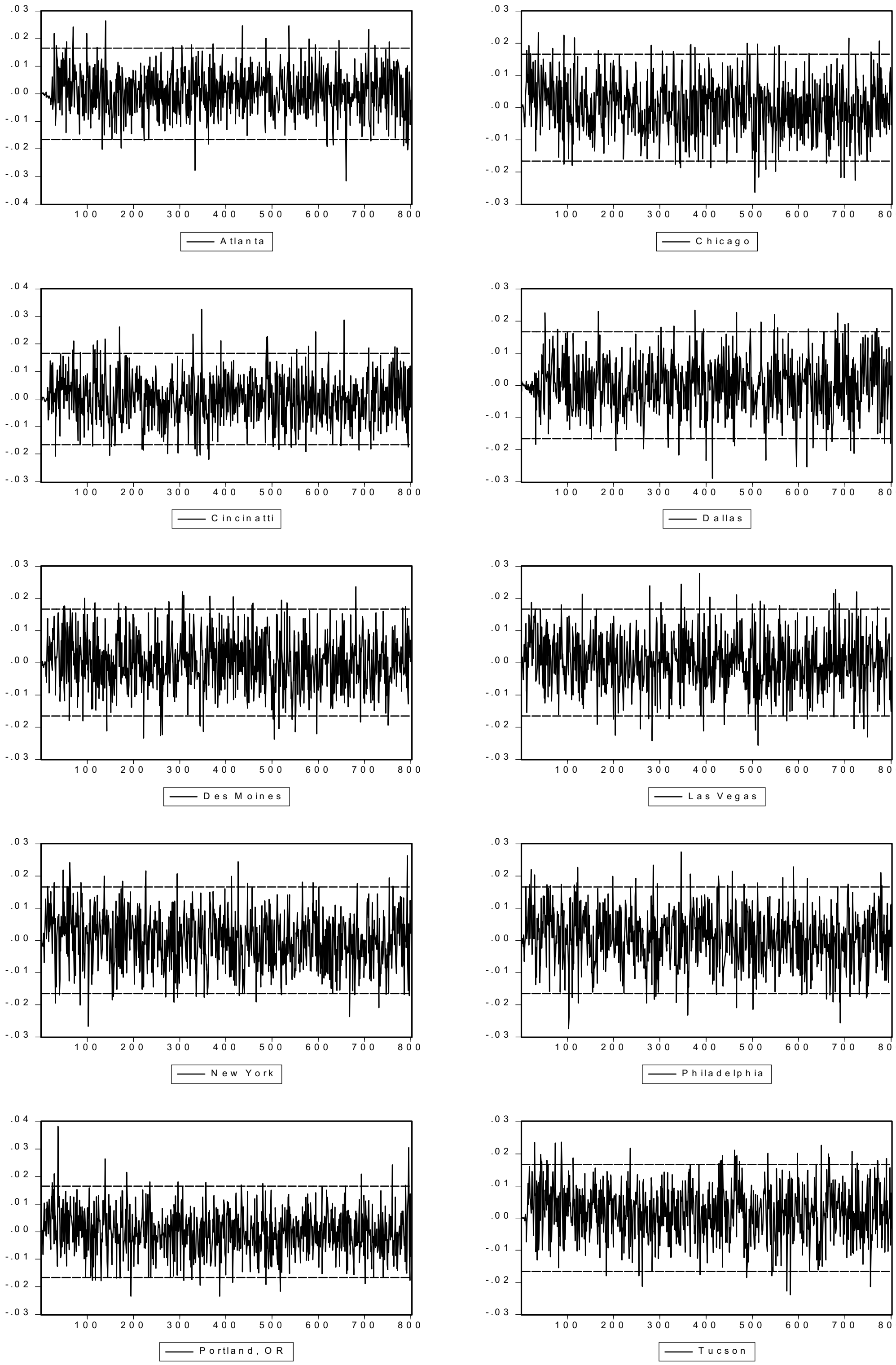


Figure 12
Correlograms of Squared Standardized Residuals

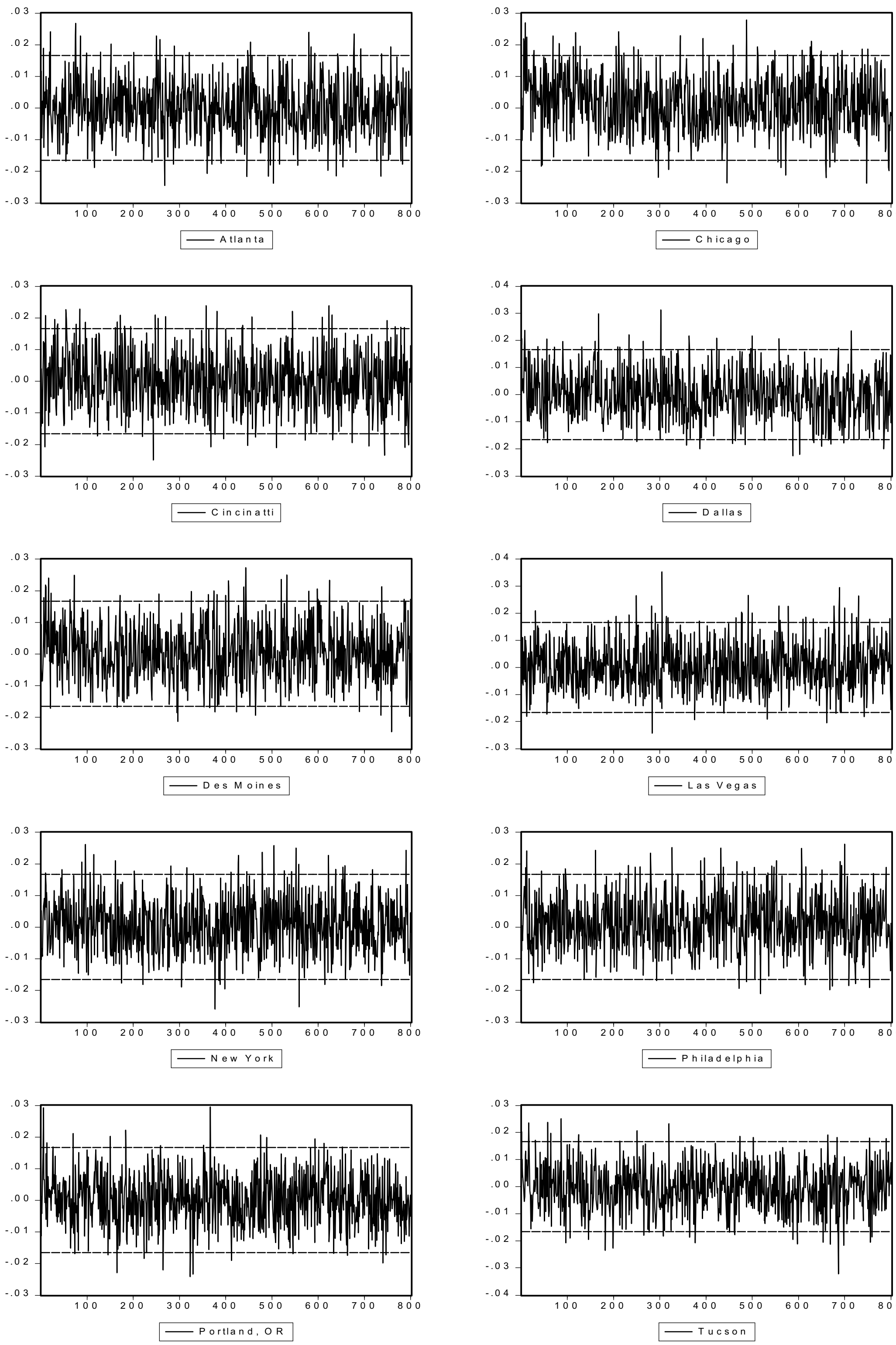
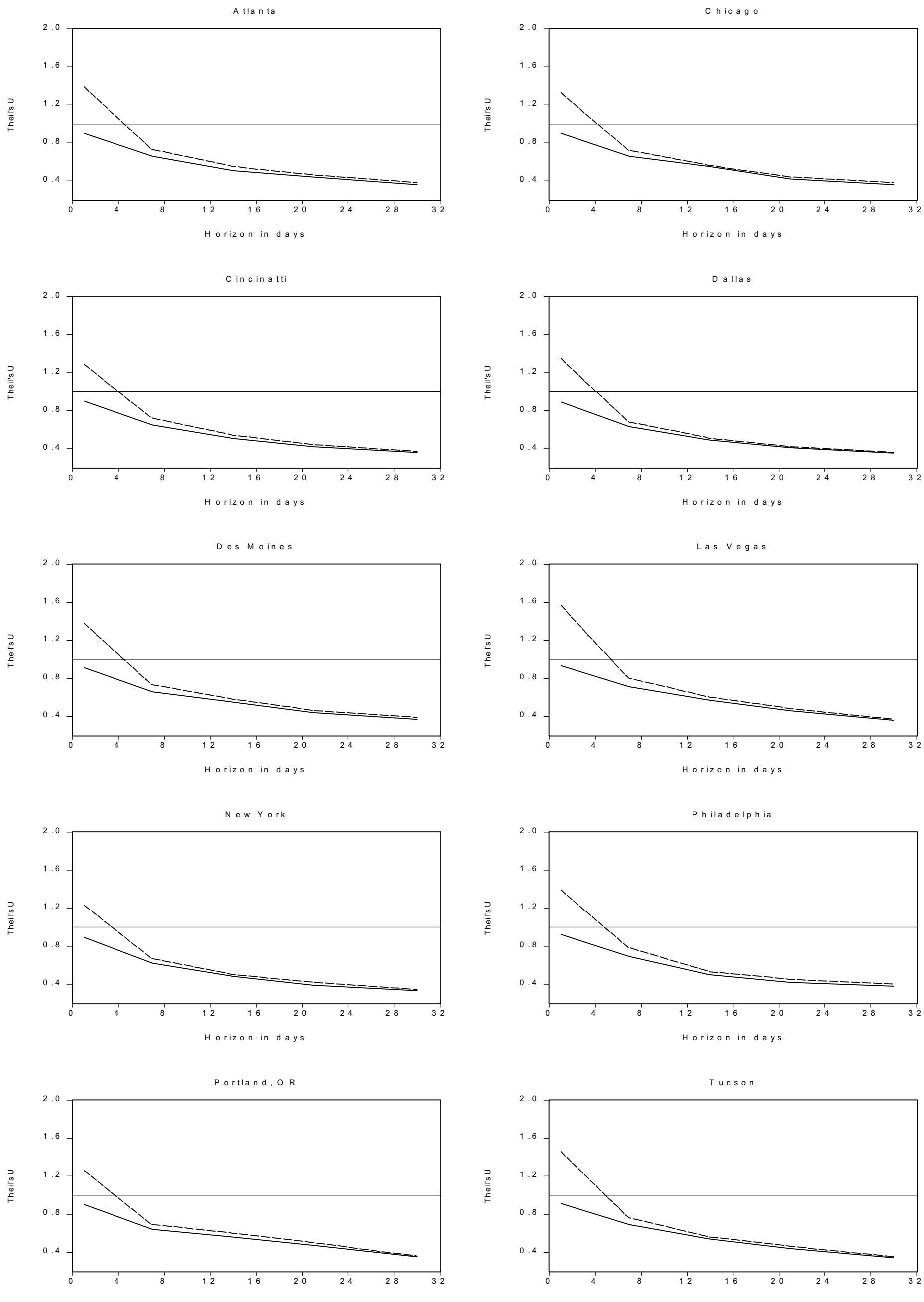
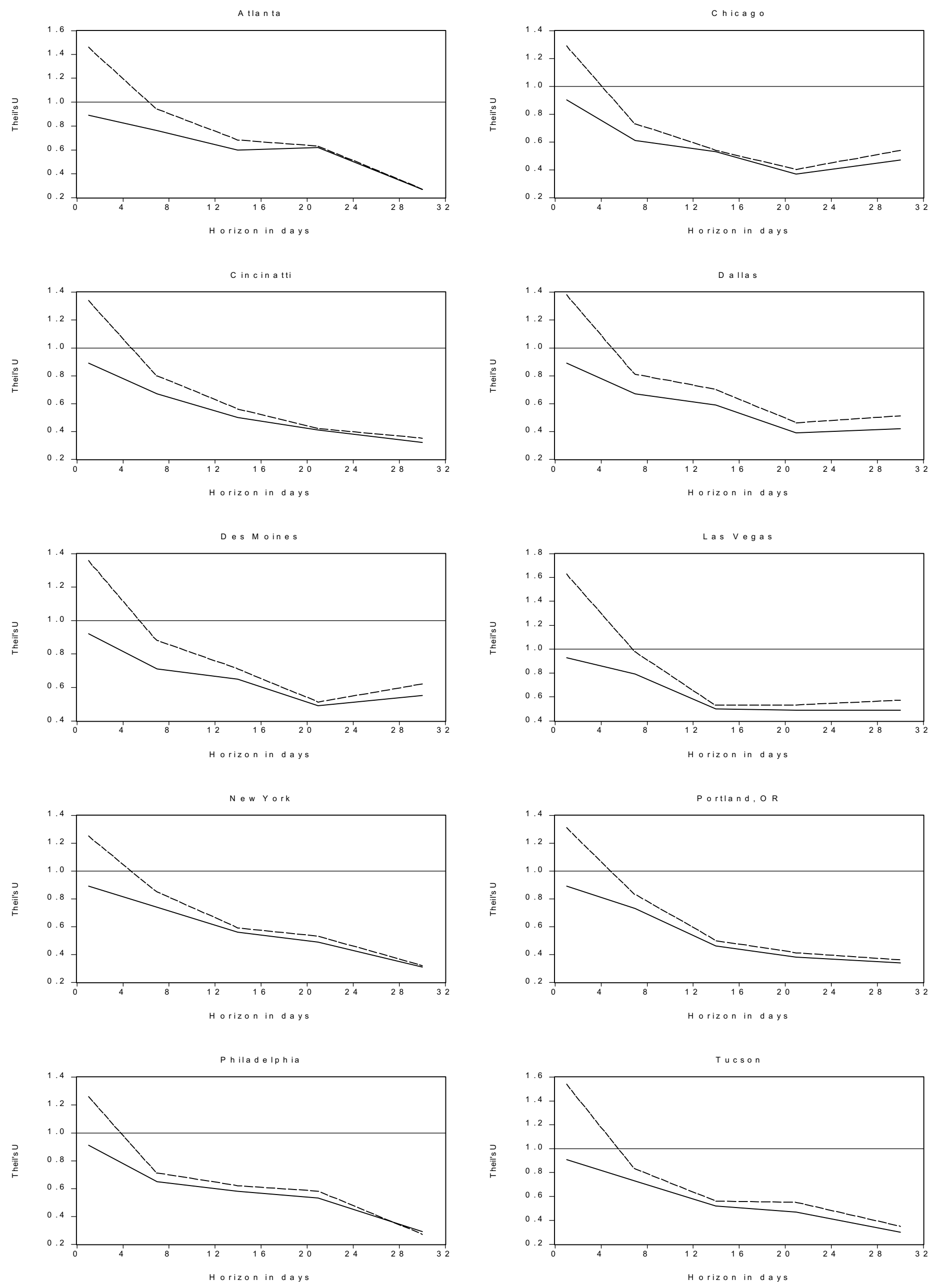


Figure 13
In-Sample U-Statistics



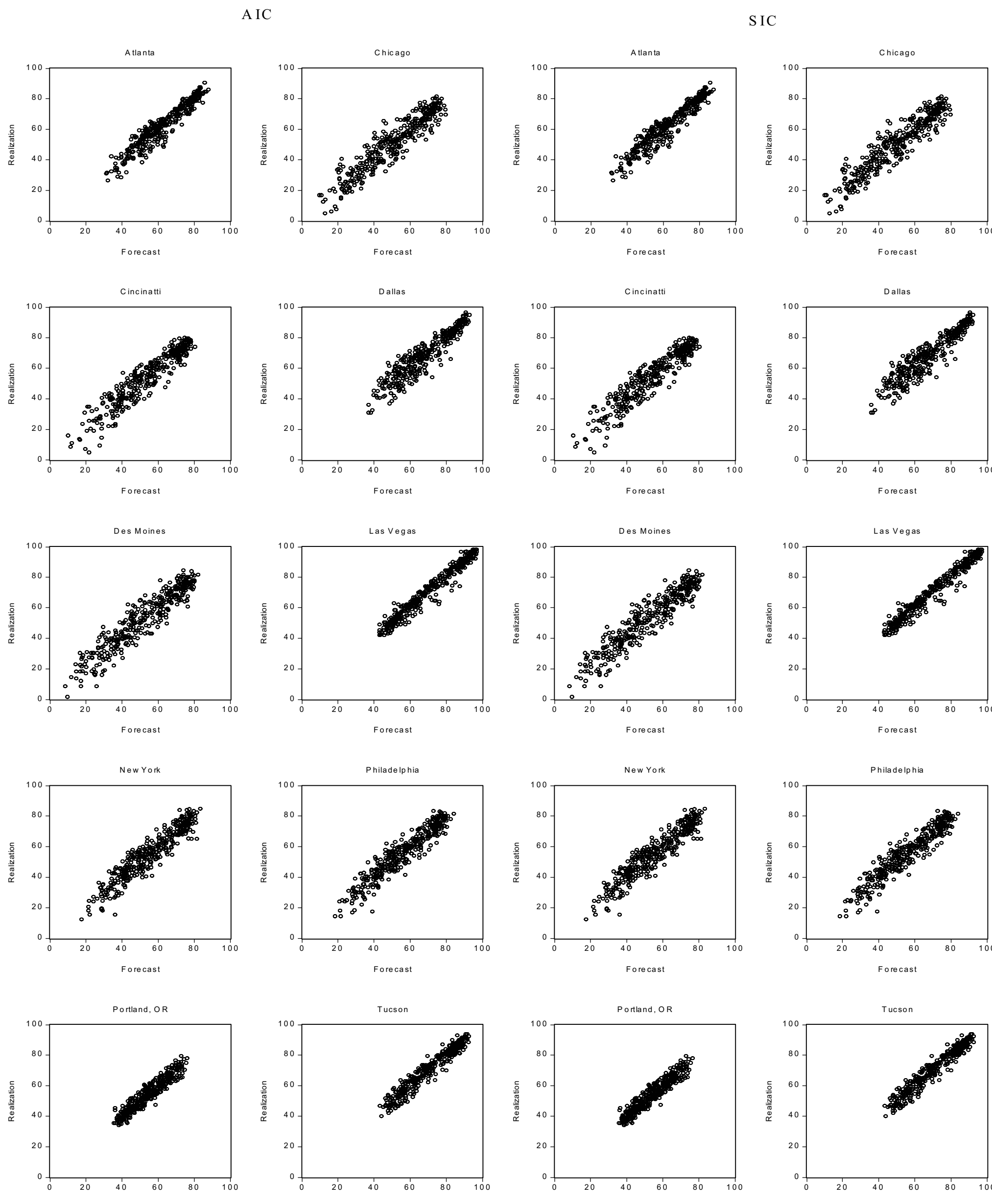
Note: Each graph displays in-sample U-statistics (the ratio of a forecast's RMSE to that of a random walk) for horizons of 1, 7, 14, 21, and 30 days. The solid line refers to a daily "seasonal+trend " model, and the dashed line refers to our AIC-selected model. For visual reference, we include a horizontal line at unity.

Figure 14
Out-of-Sample U-Statistics



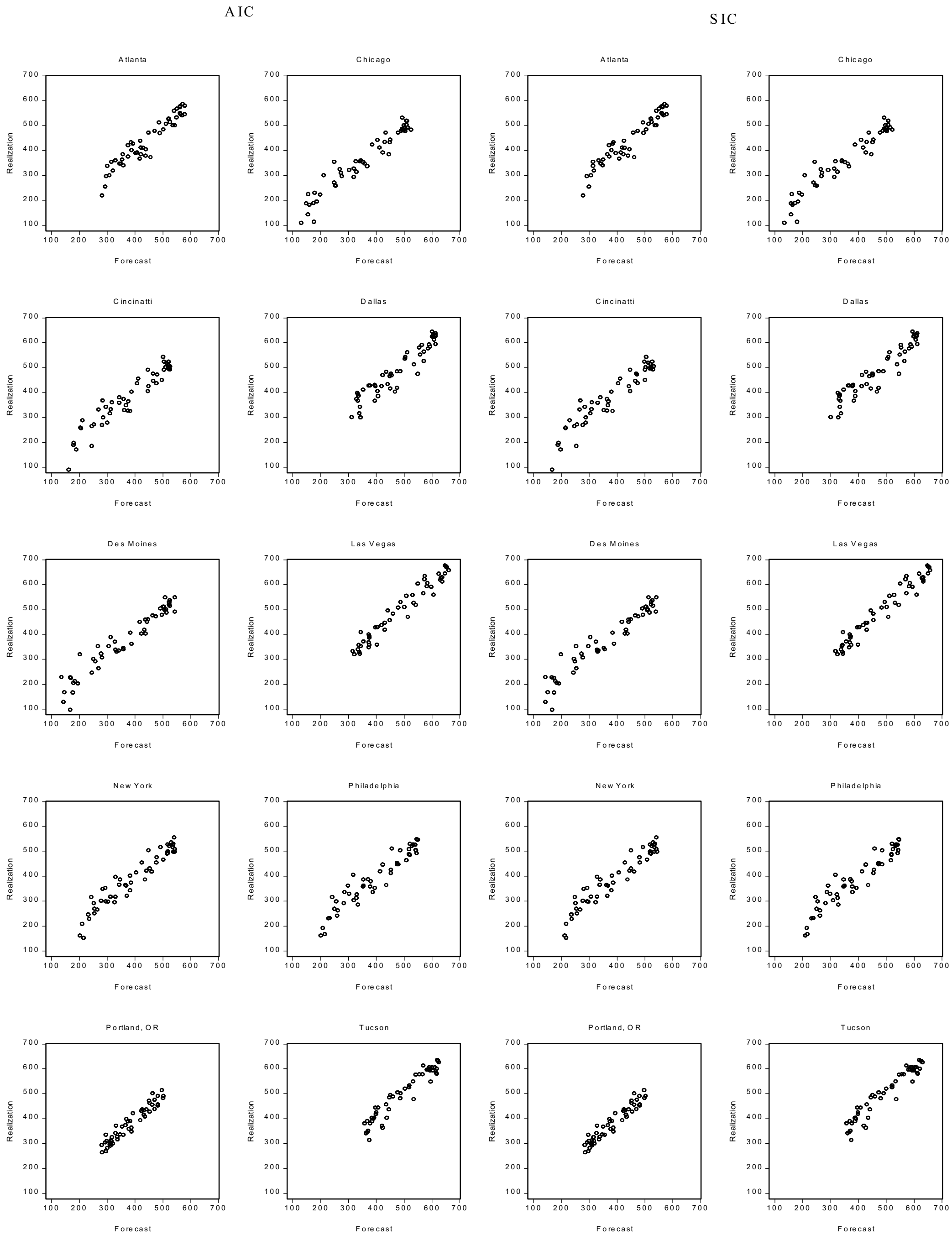
Note: Each graph displays out-of-sample U-statistics (the ratio of a forecast's RMSE to that of a random walk) for horizons of 1, 7, 14, 21, and 30 days. The solid line refers to a daily "seasonal+trend" model, and the dashed line refers to our AIC-selected model. For visual reference, we include a horizontal line at unity.

Figure 15
Daily Average Temperature Realizations and Forecasts
1-Day-Ahead



Notes: We show realizations on the vertical axis and forecasts on the horizontal axis. The two left panels are AIC-selected models and the two right panels are SIC-selected models.

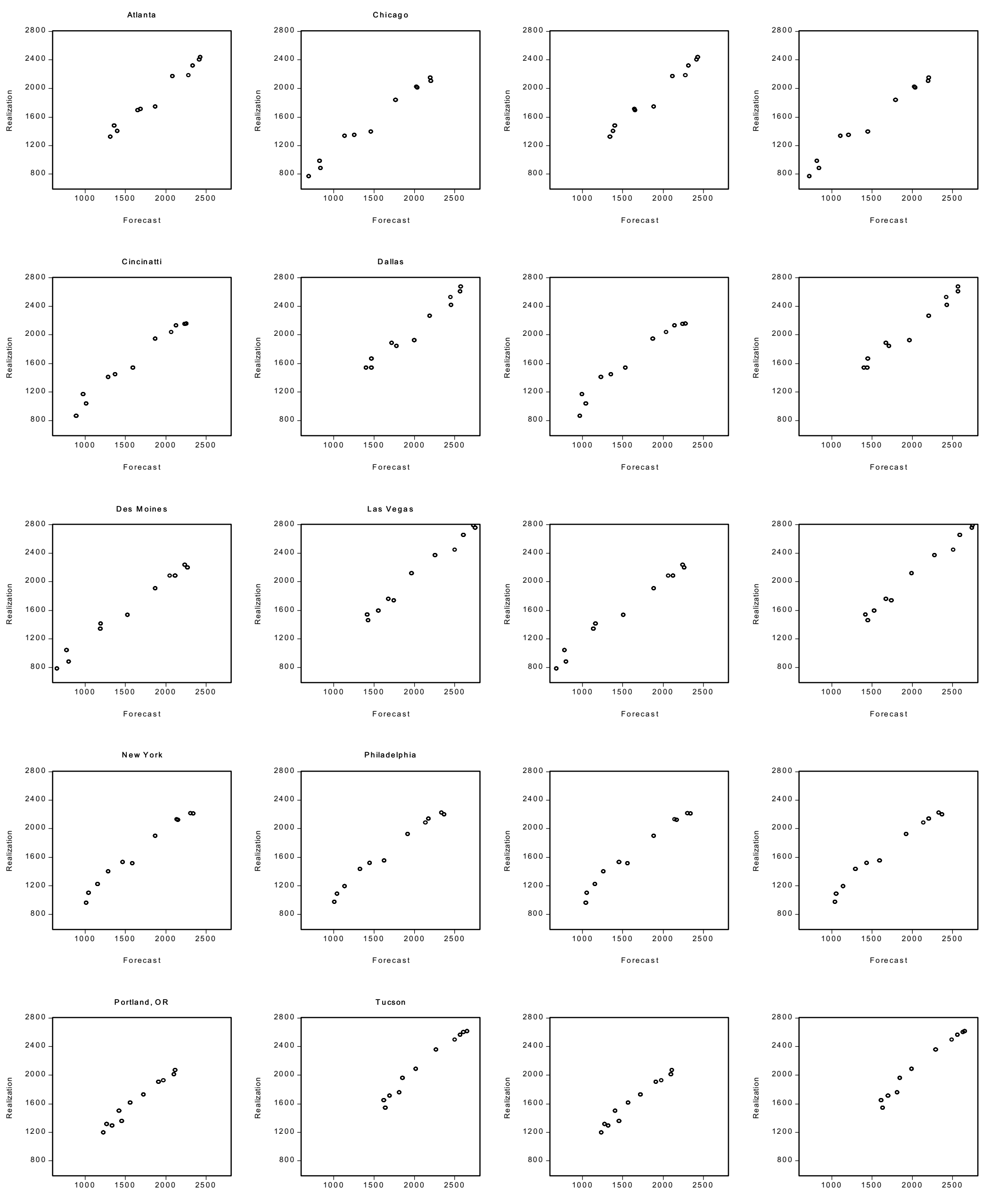
Figure 16
Daily Average Temperature Realizations and Forecasts
7-Day-Ahead Cumulative



Notes:
We show realizations on the vertical axis and forecasts on the horizontal axis. The two left panels are AIC-selected models and the two right panels are SIC-selected models.

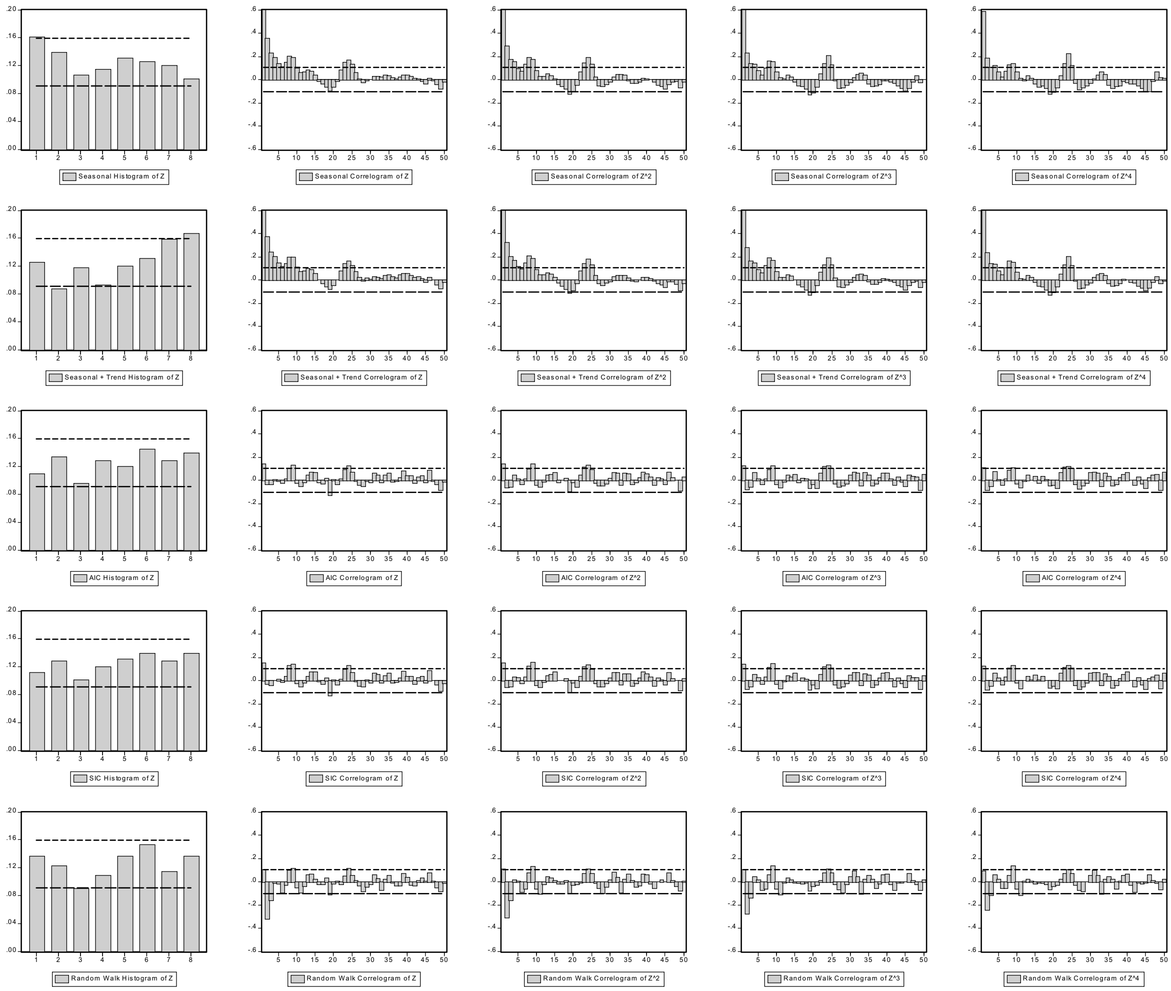
Figure 17
Daily Average Temperature Realizations and Forecasts
30-Day-Ahead Cumulative

AIC SIC



Note s:
We show realizations on the vertical axis and forecasts on the horizontal axis. The two left panels are AIC-selected models and the two right panels are SIC-selected models.

Figure 18
Adequacy of 1-Day-Ahead Density Forecast Conditional Calibration: Philadelphia



Notes: The above figure contains a histogram and the correlogram for the first 4 powers of the z series for each of the five temperature models: 1) Seasonal, 2) Seasonal + Trend, 3) AIC Selected, 4) SIC Selected and 5) Random Walk