3 Methodological Approach

3.1 Timing of Releases

The international house price database is updated quarterly, but we face great heterogeneity in the timing of each country's data releases. We have found a significant number of countries report the data needed to include a new quarterly observation of the (nominal and real) house price and PDI index series three months after the end of each quarter.[17] As a practical compromise between the timeliness of the release and the country coverage, we schedule the posting of the international house price database during four fixed periods each year incorporating already a three month lag. The schedule for a given year *T*, therefore, is:

Table 3. Yearly Timeline of Release Dates

Last Quarter Included	Release Date
Fourth Quarter of *T-1*	First full week of April of *T*
First Quarter of *T*	First full week of July of *T*
Second Quarter of *T*	First full week of October of *T*
Third Quarter of *T*	First full week of January of *T+1*

We aim to have the dataset posted during the week indicated in Table 3. We will apply the same schedule for data releases every year, unless the release lag of the country data on which the panel depends were to change, requiring us to adjust this schedule to preserve both its wide coverage and representativeness. The schedule of data releases for the current and subsequent years can be found on the web.

Whenever observations are missing for a country's house price or PDI index, we complete the series for the current release with nowcasts derived with the basic structural time series (BSTS) model that we describe in Section 3.7 below. The dataset will be revised in subsequent quarters to incorporate the country data, as the corresponding quarterly observations become available.

[17] Country coverage usually represents more than 80 percent of total output from all countries included in the sample, measured by their 2005 GDP in purchasing power parity adjusted terms according to the IMF.

3.2 Approach to Construct the Country Data: Step-by-Step

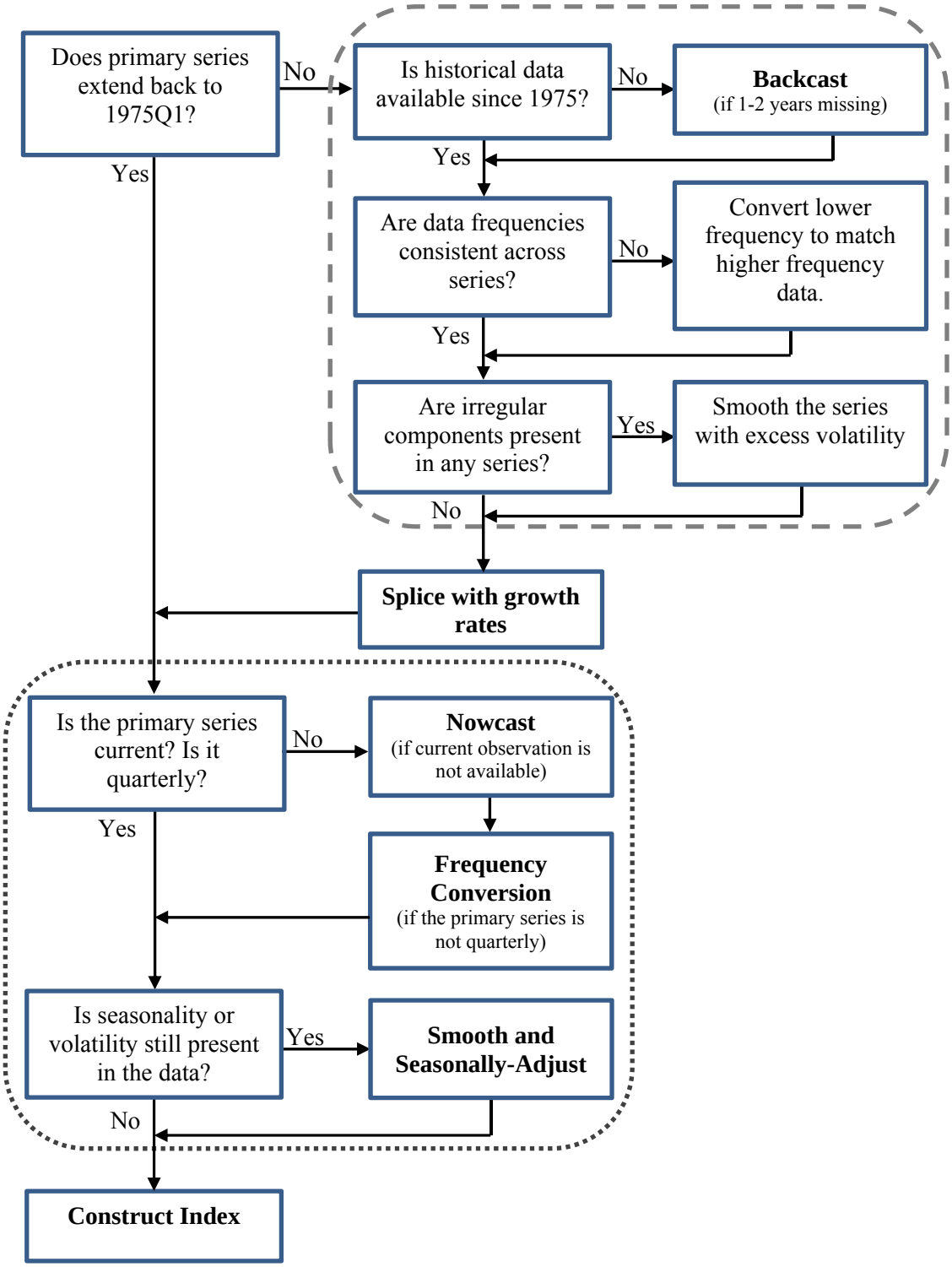
The FHFA house price index series (formerly called OFHEO house price index) serves as our benchmark measure when selecting a house price index for the other countries in the database. The main selection criteria—and preference—are given to: geographic coverage (nationwide); vintage of dwellings (existing); type of dwellings (single-family); priced unit (per dwelling); availability of data (1975 - present); and frequency (quarterly). The disposable personal income series by the Bureau of Economic Analysis (BEA) serves as our benchmark measure when selecting or reconstructing the PDI index for the other countries, although not all accounting differences across countries can be fully reconciled with the BEA concept. The PDI measure for some countries is derived from national account aggregates using the accounting relationships discussed in Section 2.2 above.

The methodological decision tree found in Figure 3 below describes the steps we follow to derive consistent series by combining the available sources for each country in the panel. The same approach is applied to all series in the international house price database. The motivation behind each step is to maintain consistency across countries, while limiting data distortions as much as possible.

The reference series for each country is its most current one, since it is selected for being the most consistent with the benchmark series (FHFA house price index and disposable personal income) and for having data available up to the present. In most cases, the reference series does not extend all the way back to the first quarter of 1975, and it has to be completed with other series which we refer to as the historical series. We label the combined series that goes back to the first quarter of 1975 as the extended or the full-length series.

Old vintages of the international house price database are kept for researchers interested in real-time data analysis.

Figure 3. Methodological Decision Tree for the Construction of Time Series Data



3.3 Frequency Conversion Methods

Frequency conversion is often required in order to report the house price and PDI indexes at a quarterly frequency. We are confronted with the problem of converting high-frequency data (monthly) to quarterly data as well as the problem of converting low-frequency data (semi-annual, annual, bi-annual) to quarterly data.

If the series has a higher frequency, then we use a simple average of the monthly observations in order to report the time series at quarterly frequency. This makes use of all available data and produces observations that summarize the quarterly patterns of the data.[18]

Interpolation is applied to the latter problem, but it is worth pointing out that interpolating procedures can only provide estimates between known data points, since quarterly data are not observed. In other words, it tries to complete the series by fitting a curve to the known data points which can then be used to infer plausible outcomes for the unknown quarterly observations. The method of interpolation will be determined by the features that are most desirable for the interpolated data. We discuss the implementation of the frequency conversion methods in greater detail in Section 3.3.1 below.

3.3.1 Quarterly Interpolation

Interpolation methods are used for the conversion of low frequency to high frequency data—that is, for temporal disaggregation whenever no additional source of high frequency data is available to facilitate the conversion.[19] Interpolation in the context of the international house price database can be defined as fitting a curve over measurements made at the sampled periods to infer unsampled quarters with which they must conform. Standard interpolating methods include constant piecewise, linear, polynomial (quadratic, cubic,…), and spline, among others.

If the observed time series has a lower frequency (semi-annual, annual, or bi-annual) than the database (quarterly), our preferred choice for interpolation is to use the quadratic-match average (or the quadratic-match sum) method to interpolate the data to a quarterly frequency.[20]

The quadratic-match methods do not guarantee that the curve fitted would satisfy non-negativity, even if the observed data points are all positive. This is a problem especially in countries that experienced periods of hyperinflation or severe inflation within the sample covered by the international house price database (such as Croatia). We resolve the issue interpolating the

[18] A note of caution, averaging may induce a spurious first-order serial correlation effect in the differenced series as shown by Working (1960) among others.

[19] Different econometric disaggregation techniques, such as the Denton (1971) and Chow and Lin (1971) approaches to cite just two of the most popular methods, can also be used for quarterization. These techniques interpolate the low frequency data at quarterly frequency using related indicator variables that are reported themselves at quarterly frequency. We generally do not have access to quarterly indicators that can be used with the available data for temporal disaggregation of the series in the international house price database.

[20] A note of caution, interpolating with a quadratic function introduces a systematic source of serial correlation in the regressors because data points are related to each other by a quadratic polynomial. This must be taken into account when using time series with interpolated data for the purpose of statistical inference—that is, standard errors should be made robust to autocorrelation in hypothesis testing.

logged series, and then use the growth rate in log-differences to splice and extend the data backwards. In all other cases, interpolation is conducted with the series expressed in levels.

Interpolation is performed in two stages. In the first stage of interpolation, the objective is to report the current and historical series at the same frequency, so they can be spliced together.[21] If the historical series is reported at a lower frequency than the current series, but the frequency of the current series is lower than quarterly, then the historical series is interpolated to the same frequency as the current series rather than at quarterly frequency. For example, in the case of Italian house prices, the historical series is reported bi-annually and the current series is reported semi-annually. In the first stage, the historical series is interpolated to a semi-annual frequency and then spliced with the current series. In the second stage of interpolation (if this step were necessary), the objective is to report the extended series at a quarterly frequency. Again, using the Italian house price series as an example, once the bi-annual series is interpolated to a semi-annual frequency and spliced with the current series, the extended series is interpolated to quarterly frequency.

3.4 Seasonal-Adjustment Method

In order for quarterly data to be useful for researchers or for policy analysis, the series must be reported to reflect the true underlying patterns of the data. Seasonal-adjustment aims to pinpoint economically relevant features of the data (trend and cycle), but irregular components (outliers, breaks) may also be present. In which case, seasonal-adjustment might not be sufficient in removing all the irrelevant noise in an observed series. Hence, we develop and implement a smoothing algorithm based on the basic structural time series (BSTS) model, fitted to the data in order to remove both seasonality and the effect of unrelated irregular noise.

We give preference to data not seasonally-adjusted by the source because then we can apply the BSTS model in a manner that consistently treats the effect of seasonality across countries and series. The seasonal adjustment procedures used by the sources often vary by country—the U.S. Census Bureau family of seasonal adjustment procedures (such as X-12) or the Bank of Spain/Eurostat TRAMO-SEATS package—which introduce heterogeneity in the data. In cases when the original data is only reported seasonally-adjusted (such as South Africa's PDI), we use it without further correction unless the seasonal adjustment by the source appears insufficient to remove all seasonality and the irregular noise in the time series. We would then apply a variant of the BSTS model to this data in order to obtain the final seasonally-adjusted, smoothed series for the international house price database.

The specification of the BSTS model in state-space form is both simple to estimate and very flexible capturing the relevant components of the data. It produces estimates of the seasonal factors that are comparable to those obtained with other conventional methods (such as X-12 and TRAMO/SEATS), while providing an integrated framework for the removal of irregular components, and can also be used for backcasting and nowcasting.[22] We further discuss the uses of the BSTS model in Section 3.7 below. For a more formal evaluation of the BSTS model for

[21] We provide details on the implementation of the splicing method in Section 3.5.
[22] We discuss the backcasting and nowcasting uses of the BSTS model in Section 3.7.

seasonal adjustment and the treatment of irregular components, the interested reader is referred to Martínez-García (2014).

If seasonality or large irregular noise components are present in the current series but not in the historical series (especially when the historical series is derived from interpolated data), we remove the seasonality and irregular components from the current series before splicing—using the BSTS model. In general, we seasonally-adjust the individual data series contained in an extended series separately whenever they display very different seasonality or irregular patterns, then splice the smoothed series. If the seasonal patterns of the individual series appear similar, they are first spliced, then seasonal adjustment is applied to the extended series. The seasonal adjustment with the BSTS model is performed on data with a frequency higher than annual, but it can be implemented with interpolated data as it removes some of the irregular patterns in the data that may follow from the implementation of the quadratic-match procedure.

3.5 Splicing Method

All series are spliced together using the growth rates of the longer historical series in order to extend the level of the shorter current time series backward in time. Splicing with exact growth rates is preferred, but we use log-differences in the case of countries that have undergone a period of severe inflation within the sample (such as Croatia). The historical time series are selected by availability in order to reasonably track the changes of the reference series.

Splicing the current series with the growth rates of the historical series may not extend the series as far back as 1975. In which case, the historical series is complemented with backcasted data, where additional observations—not representing more than two years—are obtained using the best fitted BSTS model for the available historical time series. Time series backcasting is used to extend the house price indexes of Spain and the Netherlands from the first quarter of 1976 back to the first quarter of 1975.[23] Before splicing the series, we perform the minimal adjustments needed to obtain consistent frequency across series and remove any irregular components or seasonality.

Nowcasting is applied primarily in order to ensure the timeliness and completeness of the international house price database release when the current series of reference for a country is missing some recent observations due to lags in reporting from the national sources or simply because the reported data is produced at a lower-than-quarterly frequency.

Nowcasting is used, for example, with the annual consumption of fixed capital series of Belgium and the U.K. The data at a lower-than-quarterly frequency is extended with a historical series and augmented with the nowcasted value for the unobserved year that includes the current quarter. Then, the series is interpolated to a quarterly frequency. If the current series is reported at quarterly frequency, but the reporting lag results in missing observations for the most recent quarters, we use nowcasted values for those missing observations. The nowcasted values will subsequently be replaced with actual data from their respective national sources, as those

[23] Backcasting estimates are obtained by re-ordering the original data x_t backwards from the end of the sample ($t = T$) to the beginning of the sample ($t = 1$), and then running a regression model to forecast the data of the series x_t prior to $t = 1$.

observations become available. We finally seasonally-adjust the extended quarterly series, except when the seasonal patterns or other irregular components of the data appear not to be statistically-significant.

All nowcasts and backcasts are derived from the best fitted specification of the BSTS model. This is an estimated regression model that is used for seasonal-adjustment and smoothing as well, where the underlying components of each time series—decomposed in trend, cycle, seasonal factors, and irregular noise—are fitted at each point by Maximum Likelihood and updated iteratively with the help of the Kalman filter. In Section 3.7 below, we give further interpretation to the BSTS model and a short overview of the methodology as it is implemented in the international house price database. We leave the more technical details for the interested reader to be found in Martínez-García (2014).

3.6 Approach to Aggregate the Country Data

All country series at quarterly frequency and extended going back to the first quarter of 1975 are then indexed to 2005=100. The country series are aggregated to produce global indicators of the housing market, using weights that account for the size of the economy of each country incorporated in the database. We use the 2005 purchasing power-parity adjusted (annual) GDP shares of all countries in the sample as constant weights to derive our aggregate nominal house price index and our aggregate real house price index. Similarly, we use these constant GDP weights to aggregate the nominal and real PDI series. All GDP shares are obtained from the IMF World Economic Outlook database.

3.7 More in Depth: An Overview of the Basic Structural Time Series (BSTS) Model

We use the BSTS approach as follows,

- The estimation of the BSTS model corrects for seasonality at quarterly frequency and also accounts for other extraneous irregular noise that may be present in the observed time series (historical or current). Noisy series and/or series with a statistically-significant quarterly seasonal pattern are reported as the combination of the smoothed trend and cycle components, therefore excluding both the noise and seasonality from the series.

- The BSTS model is used for backcasting and nowcasting in order to extend some time series at the beginning and end of their samples. The estimated model produces accurate backcasts/nowcasts with otherwise minimal effort in identification, but it is most appropriate either in short- to medium-horizons where observed data patterns are likely to continue or whenever changes in the time series occur slowly over time. Backcasts/nowcasts are used in practice with at most 1-2 years back/ahead for that reason.

3.7.1 All-Encompassing BSTS Model Specification

Given a time series y_t, $t = \{1,\ldots,T\}$, the standard BSTS model of Harvey (1989) decomposes the data additively as follows,

$$y_t = \mu_t + c_t + s_t + e_t,\ e_t - NID\left(0,\sigma_e^2\right),\ t = 1,2,\ldots,T, \tag{1}$$

where μ_t is the time-varying *trend*, which captures the long-run evolution of the series as a function of time; c_t is the *cycle*, which denotes transitory movements in the series around its trend; s_t represents the *seasonal factor*, which captures a recurrent movement in the series that repeats itself at fairly regular intervals; and e_t is an irregular (and random) component called *noise, error term, or disturbance term*, which includes other non-modeled, exogenous, and intrinsically irrelevant factors affecting the observed series (the signal).

These four components can also be combined multiplicatively in the time series to produce the log-additive model specification,

$$y_t = \exp\left(\mu_t\right)\exp\left(c_t\right)\exp\left(s_t\right)\exp\left(e_t\right),\ e_t - NID\left(0,\sigma_e^2\right),\ t = 1,2,\ldots,T, \tag{2}$$

which has many applications in the international house price database. This log-additive model specification reduces to the additive specification in (1) whenever we work with logged values. Hence, the BSTS model can be discussed solely in its additive form even though we actually consider the log-additive case as well in our implementation.

All four constituent components (μ_t, c_t, s_t, e_t) are stochastic, and the disturbances driving them are mutually uncorrelated in general. The extraneous noise component e_t is modeled as white noise, while the trend μ_t, cycle c_t and seasonal factor s_t are all assumed to follow a particular model. As stated earlier, we consider a stochastic polynomial model of order two or lower for the trend component μ_t. We assume a covariance-stationary AR(p) process for the transitory component of the time series described by the cycle c_t. We also assume a stochastic version of the standard "seasonal factors" model adding up to a random shock to capture time-variation in the seasonality. For the reasons stated before, we consider both the additive as well as the log-additive specifications of the BSTS model under consideration but describe the modeling choices for the unobserved components (μ_t, c_t, s_t) on the additive case only. For more details on the exact state-space form of each variant of the BSTS model that we actually explore and a further discussion of the methodology and its advantages, the interested reader is referred to Martínez-García (2014).

The additive, univariate specification of the <u>all-encompassing BSTS model</u> is given by,

$$\forall t = \{1, ..., T\}$$

$$\begin{cases} y_t = \mu_t + c_t + s_t + e_t,\ e_t - NID\left(0, \sigma_e^2\right), & \text{signal equation (decomposition equation),} \\ \\ \mu_t = \mu_{t-1} + b_{t-1} + \varepsilon_t,\ \varepsilon_t - NID\left(0, \sigma_\varepsilon^2\right), & \text{state equation (local polynominal trend),} \\ b_t = b_{t-1} + \delta_{t-1} + \xi_t,\ \xi_t - NID\left(0, \sigma_\xi^2\right), \\ \delta_t = \delta_{t-1} + \zeta_t,\ \zeta_t - NID\left(0, \sigma_\zeta^2\right), \\ \\ c_t = \phi_1 c_{t-1} + \phi_2 c_{t-1}^1 + \vartheta_t,\ \vartheta_t - NID\left(0, \sigma_\vartheta^2\right), & \text{state equation (AR(2) cycle),} \\ c_t^1 = c_{t-1}, \\ \\ s_t = -s_{t-1} - s_{t-1}^1 - s_{t-1}^2 + \omega_t,\ \omega_t - NID\left(0, \sigma_\omega^2\right), & \text{state equation (stochastic seasonal),} \\ s_{t+1}^1 = s_t, \\ s_{t+1}^2 = s_t^1, \end{cases} \tag{3}$$

where the component μ_t represents the unobserved level, b_t determines the unobserved linear term on the time-trend, and δ_t is the unobserved quadratic term on the time trend. For specifications with a cyclical component c_t, an AR(2) (or AR(1)) process is used for reference. For specifications with seasonality, the quarterly seasonal component s_t is recurrent and its time-

invariant coefficients must sum up to a random disturbance. The random disturbances e_t, ε_j, ξ_t, ζ_t, ϑ_t, and ω_t are normally distributed, and mutually independent at all leads and lags.

The parameters of the model include up to six variances for the random disturbances, $\sigma_e^2 > 0$, $\sigma_\varepsilon^2 > 0$, $\sigma_\xi^2 > 0$, $\sigma_\zeta^2 > 0$, $\sigma_\vartheta^2 > 0$, and $\sigma_\omega^2 > 0$. The list of parameters also includes *p* (*p≤2* in our specification of reference) autoregressive coefficients used to describe the AR(p) cyclical component of the time series. The model description is incomplete without a specification of priors for $\left(\mu_0, b_0, \delta_0, c_0, c_0^1, s_0, s_0^1, s_0^2\right)$, which we generally take to be diffuse priors in all our estimations.

3.7.2 BSTS Model Estimation, Backcasting/Nowcasting and Forecasting

We represent all variants of the BSTS model in state-space form and estimate their time-varying, unobserved components by Maximum Likelihood. The advantage of assuming the normality of the disturbances is that it simplifies the estimation of the model in state-space form by Maximum Likelihood, and also allows us to use the Gaussian Kalman (Bucy) filter. The Kalman filter algorithm is applied to recursively update the model with new observed data that contains noise (random variation) to produce a statistically optimal prediction and forecast of the unobserved states (cycle, trend, and seasonal) that characterizes the time series along with its corresponding uncertainties. The main assumptions of the Kalman filter are that the model be a linear dynamic system and that the error term (irregular noise on the observed series) and all state disturbances have a Gaussian distribution.

Nowcasting and forecasting are fairly standard in this case. Backcasting is accomplished by re-ordering the original data y_t from $t=T$ to $t=1$ (in reverse order) and then running the estimation and forecasting algorithm under the same state-space model specification chosen to fit the data. Maximum Likelihood estimation and the recursive Kalman filter are exploited to obtain estimates of the initial states and forecasts of the series in reverse order (backcasts).

3.7.3 BSTS Model Selection

Selecting an appropriate model specification depends on: (a) whether the data is trending or not (in other words, the series is non-stationary)? (b) whether the series has a transitory component or cycle that is stationary? (c) whether there is seasonality in the data? and (d) whether the noise and irregular components are well-approximated as independent and identically distributed Gaussian random disturbances? Model selection would consider alternative specifications of the cyclical AR(p) component and the seasonal model—including variants without them—based on standard likelihood-based selection criteria (primarily the Akaike Information Criterion), strongly favoring the less-parameterized and less-complex specification of the model. For further discussion on the goodness-of-fit in-sample and the forecast accuracy out-of-sample as well as the actual implementation of the model selection procedures, the interested reader is referred to Martínez-García (2014).